

OM of: US-09-471-276-831 to: EST:* out-format: pfs
Date: Jan 7, 2002 5:20 PM

About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=firmer-p2n.model -DEV=xih
-O=cgrr2.1/USPto.spool/US09471276/runat_07012002_142444_19876/app_query.fasta_1.274
-DB=EST -CPMT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.050 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=6.500
-GAPEXT=7.000 -YGAPOP=10.000 -YGAPOP=10.000 -DELCP=6.000
-DELCP=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.ccd
-LIST=100 -DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0
-ALIGN=45 -MAXLEN=200000000 -USER=US09471276.ecgrr1_1.3762
-MINLEN=0 -MAXLEN=200000000 -USER=US09471276.ecgrr1_1.3762
-NCPU=6 -ICPU=3 -LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-471-276-831
Query length: 126

Database: EST:*

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 1532.110000

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_est1:AL531425	+	573.00	1019.45	1.3e-47	612	AL531425 LTI.NFL001.NH
gb_est1:W250599	+	496.00	884.39	4.3e-40	457	W250599 Zeb68b07.r1 Soares.fetcl
gb_est1:AA250460	+	125.00	220.45	0.0041	455	AA250460 mw99e06.r1 Soares.fetcl
gb_est1:AA700860	-	114.00	213.74	0.0097	115	AA700860 z136b07.s1 Soares.fetcl
gb_est1:BG397331	+	107.50	183.70	0.4551	916	BG397331 60243123P1 NIH.MGC.48
gb_est1:AA530090	+	105.00	184.50	0.4108	506	AA530090 v137c01.r1 Stratagene.liv
gb_est1:HG706583	+	104.00	182.29	0.5458	531	HG706583 60267296P2 NIH.MGC.96
gb_est1:AM395511	+	92.50	167.80	3.50	268	AM395511 46143 MARC.2P1G Sus.sc
gb_est1:AM456421	+	92.50	167.63	3.57	273	AM456421 76621 MARC.2P1G Sus.sc
gb_est1:BI337778	+	92.50	162.02	7.35	514	BI337778 361317 MARC.1P1G Sus.s
gb_est1:BE012583	+	92.50	161.51	7.84	544	BE012583 122203 MARC.1P1G Sus.s
gb_est1:BG733159	+	92.50	161.45	7.90	548	BG733159 346903 MARC.1P1G Sus.s
gb_est1:TB68210	+	92.00	175.92	1.23	97	TB68210 y440b03.r1 Stratagene.liv
gb_est1:BG819104	-	91.00	155.31	17.37	809	BG819104 60278119P1 NCI.CGAP.B
gb_est1:BG775650	-	90.00	152.16	26.00	943	BG775650 602650592P1 NIH.MGC.40
gb_est1:BF525381	-	90.00	150.78	31.02	1101	BF525381 60206945P1 NCI.CGAP
gb_gss:AO400188	-	89.00	156.00	15.88	500	AO400188 m9b000612P2 CUGI.R1C6
gb_est1:AM056411	-	88.50	151.89	7.47	233	AM056411 prlu-40 Representation
gb_est1:AM656229	-	88.50	154.61	18.99	539	AM656229 108362 MARC.1BOV Bos.t
gb_est1:BG178413	+	87.50	148.16	43.30	852	BG178413 60233020P1 NIH.MGC.91
gb_est1:AL532540	+	87.50	147.91	44.85	920	AL532540 AL532540 LTI.NFL001.NH
gb_est1:BF441074	+	87.00	152.98	23.41	470	BF441074 257194 MARC.2P1G Sus.s
gb_gss:TA319D090	+	87.00	151.49	28.35	556	TA319D090 T. brucei shared gene
gb_gss:AO939939	+	87.00	150.84	30.80	598	AO939939 Sheared DNA-13C17.TP.S
gb_est1:BG382087	+	86.50	151.43	28.56	506	BG382087 297863 MARC.1P1G Sus.s
gb_est1:BG609827	+	86.50	150.72	31.27	548	BG609827 323811 MARC.1P1G Sus.s
gb_est1:BG350746	+	86.00	145.68	50.25	874	BG350746 60243026P1 NIH.MGC.18
gb_est1:BG355440	+	85.00	145.61	60.67	797	BG355440 HVM610002L07E Hordeum
gb_est1:BG337700	+	85.00	143.20	82.03	945	BG337700 60243382P1 NIH.MGC.40
gb_est1:BG369925	-	84.50	142.88	85.48	886	BG369925 60243382P1 NIH.MGC.26
gb_est1:BE542780	+	84.00	140.69	113.17	1025	BE542780 60106638P1 NIH.MGC.1
gb_est1:AM001152	+	83.50	145.52	55.63	481	AM001152 wu5c03.x1 Soares.Dieq
gb_est1:BE753233	+	83.50	145.13	63.03	562	BE753233 206168 MARC.2BOV Bos.t
gb_gss:CN502DVB	-	83.00	141.52	101.73	844	AL233600 Tetradion n19r01r1dis
gb_est1:BF606165	-	83.00	146.97	50.56	413	BF606165 273126 MARC.3BOV Bos.t
gb_est1:BE753245	+	82.50	143.78	76.14	535	BE753245 206184 MARC.2BOV Bos.t
gb_est1:BG442106	+	82.50	140.30	119.01	792	BG442106 GA.E0015M19F Gossyp
gb_est1:BG177264	-	82.50	139.19	137.30	888	BG177264 60231445P1 NIH.MGC.85
gb_est1:AL535624	-	82.00	142.01	99.63	591	AL535624 AL535624 LTI.NFL006.PI
gb_est1:AL110419	+	82.00	139.77	127.34	760	AL110419 DKFZPA344L1031.r1 434

gb_est1:BI331839	-	82.00	138.09	158.09	919	BI331839 60298251P1 NCI.CGA
gb_est1:BF347835	-	82.00	136.48	194.19	1101	BF347835 60202297P1 NCI.CGA
gb_est1:AL597254	+	81.50	144.17	72.50	419	AL597254 DKFZPA338P1 NCI.CGA
gb_est1:BG257260	+	81.50	137.76	164.83	862	BG257260 60237797P1 NIH.MGC
gb_est1:BF155074	+	81.00	140.37	117.96	581	BF155074 7694612.x1 NCI.CGAP
gb_est1:BG389779	-	81.00	138.56	148.69	712	BG389779 60241515P1 NIH.MGC
gb_est1:BG170904	+	81.00	137.54	169.55	799	BG170904 60232667P1 NIH.MGC
gb_est1:BG831932	+	81.00	136.41	196.12	908	BG831932 60276559P1 NIH.MGC
gb_est1:WV66493	+	80.50	140.09	122.23	542	WV66493 AV66493 Bos.taurus
gb_est1:AV667180	+	80.50	140.06	122.74	544	AV667180 AV667180 Bos.taurus
gb_gss:AO036774	-	80.50	139.65	129.45	570	AO036774 CIT-HSP-233C18.TF
gb_est1:BG948864	+	80.50	139.40	133.59	586	BG948864 OVI-HB0032-300101-5
gb_est1:BE951615	+	80.50	137.65	167.29	714	BE951615 OVI-M-CCO-9-03-0
gb_est1:BG477108	+	80.50	136.92	181.65	775	BG477108 60252430P1 NIH.MGC
gb_est1:BF342722	+	80.50	135.83	211.14	876	BF342722 60190288P1 NIH.MGC
gb_est1:BE547838	+	80.50	135.55	218.84	904	BE547838 60107472P1 NIH.MGC
gb_est1:BG393190	+	80.50	132.40	328.04	1290	BG393190 60241160P1 NIH.MG
gb_est1:AK013768	-	80.50	128.52	539.21	1996	AK013768 Mus musculus adult
gb_est1:BG669105	-	80.00	140.26	119.66	481	BG669105 336989 BANC.5BOV Bo
gb_gss:AO615445	-	80.00	139.45	132.77	527	AO615445 HS.5144.B1.D04.77A
gb_est1:BE655825	-	80.00	139.09	139.10	549	BE655825 155005 MARC.ABOV Bo
gb_est1:BP040860	+	80.00	138.11	157.70	613	BP040860 U1-R-B2-Bet-C-04-0
gb_est1:BF204121	+	80.00	136.58	191.80	728	BF204121 60186906P1 NIH.MGC
gb_est1:BF526738	+	80.00	136.53	193.00	732	BF526738 60207057P1 NCI.CGA
gb_est1:BF397882	+	80.00	134.78	241.71	892	BF397882 60203498P1 NCI.CGA
gb_est1:BG849292	+	79.50	132.08	316.84	1093	BG849292 60279354P1 NIH.CG
gb_est1:BG979577	+	79.00	140.99	108.89	363	BG979577 CM4-CNO062-120101-7
gb_est1:AM873567	+	79.00	138.29	154.06	491	AM873567 OH7BD10.s1 NCI.CGAP
gb_gss:AO374839	+	79.00	138.13	157.28	500	AO374839 RPI111-14B115.TV.RP
gb_est1:AI231406	+	79.00	138.09	157.99	502	AI231406 EST228094 Normalized
gb_est1:BF668565	+	79.00	133.23	294.71	868	BF668565 60212361P1 NIH.MGC
gb_est1:BG417958	+	79.00	132.79	311.77	912	BG417958 HVM600200C15F.Hord
gb_est1:BF072923	+	78.50	138.15	156.84	451	BF072923 219433 MARC.2BOV Bo
gb_gss:AA297757	+	78.50	137.95	160.80	461	AA297757 RPI1-23-10169.TV.RP
gb_est1:BE969690	+	78.50	135.20	228.05	629	BE969690 60166091P1 NIH.MGC
gb_est1:BI104222	-	78.50	133.09	299.91	797	BI104222 60288977P1 NCI.CGA
gb_est1:BG688884	+	78.50	132.76	313.22	828	BG688884 60270329P1 NCI.CGA
gb_est1:BF232983	+	78.50	132.47	324.88	855	BF232983 60202350P1 NCI.CGA
gb_est1:BE902495	+	78.50	132.01	344.85	901	BE902495 60167578P1 NIH.MGC
gb_est1:BG247598	-	78.00	131.37	374.19	968	BG247598 60235927P1 NCI.CGA
gb_est1:AU121804	+	78.00	132.22	335.37	795	AU121804 AU121804 MAMMAL Hom
gb_est1:R33896	+	77.50	138.49	150.21	335	R33896 EST59761 Human Brain
gb_est1:R90357	+	77.50	136.87	184.86	426	R90357 yu96d08.r1 Soares.fet
gb_gss:AO941634	+	77.50	134.65	245.72	547	AO941634 Sheared DNA-43B10.T
gb_gss:AO005559	-	77.50	134.41	253.41	562	AO005559 MR3-GN018F-241100-0
gb_gss:AO084301	-	77.50	133.98	267.83	590	AO084301 RPI1-23-35H12.TV.RP
gb_gss:AO020851	-	77.50	133.56	282.35	618	AO020851 CIT-HSP-2310J3.TP.C
gb_est1:BI115008	+	77.50	132.16	338.11	724	BI115008 60286284P1 NIH.MGC
gb_gss:AC015261	+	77.50	131.98	346.10	739	AC015261 Homo sapiens genome1
gb_est1:AG509621	+	77.50	131.03	390.69	822	AG509621 60186154P1 NIH.MGC
gb_est1:AL549756	+	77.00	138.85	143.30	308	AL549756 AL549756 LTI.NFL006
gb_est1:BP059333	+	77.00	134.27	257.86	516	BP059333 7K60E05.x1 NCI.CGAP
gb_est1:AS574120	+	77.00	134.09	264.13	527	AS574120 ng51e04.s1 NCI.CGAP
gb_est1:AM656141	+	77.00	133.63	280.17	555	AM656141 108224 MARC.1BOV Bo
gb_gss:AO573775	-	77.00	129.08	501.79	926	AO573775 nbx00844H16F CUGI.R

seq_name: gb_est1:AL531425
seq_documentation_block: 612 bp mRNA
LOCUS AL531425 LTI.NFL001.NBC4 Homo sapiens cDNA clone CSDDM002YF08 5
DEFINITION prime, mRNA sequence.
ACCESSION AL531425
VERSION AL531425.1 GI:12794918
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 612)
 AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1. 612
 /organism:"Homo sapiens"
 /db_xref:"taxon:9606"
 /clone:"CSODM002YF08"
 /clone_lib:"LIT_NFL001_NBC4"
 /sex:"male"
 /library_type:"neuroblastoma cells"
 /lab_host:"DH10B"
 /note:"Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : filiang@life.com URL : http://fulllength.invitrogen.com"

BASE COUNT 112 a 187 c 189 g 121 t 3 others

ORIGIN

alignment_scores: Quality: 573.00 Length: 127
 Ratio: 4.697 Gaps: 1
 Percent Similarity: 96.063 Percent Identity: 95.276

alignment_block:
 US-09-471-276-831 x AL531425 ..

Align seg 1/1 to: AL531425 from: 1 to: 612

```

1 MetSerMetLeuValValPheLeuLeuTrpGlyValThrTrpGlyPr 17
|||||
28 ATGTCATGCTCGTGGCTTCTCTGCTGCGGTGTACACGCGGCC 77
|||||
17 ovalThrGluAlaAlaIlePheTyrGluThrGln**SerLeuTrpAlaG 34
|||||
78 AGTCACAGACGACGCCATATTTATGAGAMGACGCCACGCTCGGCAG 127
|||||
34 luserGluHis**LeuLysThrLeuGlnGlnGlnGlnGlnGlnGlnGln 50
|||||
128 AGTCGCAATCATGCTGTA. ACCCTTGCCCAATGTCAGCTGACGTC 176
|||||
51 GTPProGlyAspSerArgLeuProAlaValGlnGlnGlnGlnGlnGln 67
|||||
177 GCMCGCGCTGAGAGCTCAGACTTCGCGCTTCAACATGAGGGTGGCCC 226
|||||
67 lnguProValHisLeuAspSerProAlaIleLysHisGlnPheLeu 83
|||||
227 ACGAGCTGTGACACTGACTGACTCCATCAACGACACGATTCCTGCTG 276
|||||
84 ThrGlyAspThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
|||||
277 ACGGCTGACACCCAGGCGCGCTACCGCTGCGCGCTGCTGTCACAGG 326
|||||
100 YTPP**GlnLeuSerLysLeuLeuGlnGlnGlnGlnGlnGlnGlnGln 117
|||||
327 ATGACACCTGAGCAAGCTCTGAGAGCTGACAGGCGCAAGTCTTG 375
|||||
117 lacySerLeuAlaLeuAspGlyAlaSer 126
|||||
376 CTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 404

```

seq_name: gb_est2:W25099

seq_documentation_block:
 LOCUS W25099 457 bp mRNA EST 20-AUG-1996
 DEFINITION zb68b07.r1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
 IMAGE:308725.5, similar to SW:ALIG_HUMAN P04217
 ALPHA-1B-GLYCOPROTEIN. [1] ;, mRNA sequence.

ACCESSION W25099
 VERSION W25099.1 GI:1302954
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 457)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Donald, M.F., Chiapelli, B.,
 Chissole, S., Dietrich, N., Dubuque, T., Favelli, A., Gish, M., Hawkins,
 B., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
 Schellander, G., Soares, J., Prange, C., Riffkin, L., Rohlfing, T.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1951 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 267.
 Location/Qualifiers
 1. 457
 /organism:"Homo sapiens"
 /db_xref:"GDB:1252138"
 /db_xref:"taxon:9606"
 /clone:"IMAGE:308725"
 /clone_lib:"Soares_fetal_lung_NbHL19W"
 /dev_stage:"19 weeks"
 /lab_host:"DH10B (ampicillin resistant)"
 /note:"Organ: lung; Vector: pRT73D (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer
 [5'-GTTCACCAATCTGAGAGCGGCGCGCAATTTTGTGTGT-3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pRT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot "5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. This library was constructed
 from the same fetus as the fetal heart library, Soares
 fetal heart NbHL19W."

BASE COUNT 90 a 136 c 129 g 91 t 11 others

ORIGIN

alignment_scores: Quality: 496.00 Length: 128
 Ratio: 4.313 Gaps: 4
 Percent Similarity: 89.844 Percent Identity: 87.500

alignment_block:
 US-09-471-276-831 x W25099 ..

Align seg 1/1 to: W25099 from: 1 to: 457

```

1 MetSerMetLeuValValPheLeuLeuTrpGlyValThrTrpGlyPr 17
|||||

```


normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGI, National Institutes of Health). Note: this is a NIH_MGC library.*

BASE COUNT 116 a 173 c 151 g 91 t
ORIGIN

alignment_scores:
Quality: 104.00 Length: 116
Ratio: 1.763 Gaps: 16
Percent Similarity: 50.862 Percent Identity: 34.483

alignment_block:
US-09-471-276-831 x BG706583 ..

Align seg 1/1 to: BG706583 from: 1 to: 531

```

2 SerMetLeuValValPhe.....LeuLeuLeuTrpGluValThrTrpG1 16
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
37 GCCATGATCCGACCTTCACGGCTCTCTCTCCCTCGCGCTGAGTCTGGG 86
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
16 yProValThrGluAlaAlaIlePheTyTGluThrGln***SerLeuTrpA 33
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
87 CCCGAGGACCGACATCGAGCGAGGCCCTCCCAACACCACTCTCTGGG 136
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
33 LeuGluSerGluHis**LeuLysThrLeuGluGln...CysAspAlaAsp 48
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
137 CTGAGCGCAGGCTCTGTGATCAG...CTGGGCGAAGCTGTGACCACTCTG 182
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
49 ValProGluProProGluYAspSerArgLeuPro..... 59
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
183 GTTCACAGCGGACCTCGAGGCTCGGAGTACCTCTGATACAGAGAGAC 232
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
60 .AlaValGlnGluTrpGluAlaGlnGluProValHisLeuAspSerProA 76
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
233 AAGCCGACGACCCCTGGGAGACAGACAGACCACTG.....GAGCCCA 273
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
76 IaIleLysHisGlnPhe.....LeuLeuThrGluAspThrGlnGly 89
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
274 AGAACAAGCGGACGATCTCCATCCCATGACAGAGAGACACTATGACAGG 323
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
90 ArgTyArgCysArgSerGlyLeuSerThrGlyTrp**GlnLeuSer 105
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
324 AGATACGCGGTGTACTTCCAGCCCTGTAGCGCTGTGACAGAGACACT 371
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

seq_name: gb_est1:AM359511

seq_documentation_block:
LOCUS AM359511 268 bp mRNA EST 09-JUL-2000
DEFINITION 46143 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AM359511
VERSION AM359511.1 GI:6864161
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 268)
AUTHORS Fahnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W., and Keele,J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine

JOURNAL COMMENT Unpublished (2000)

CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred

v0.980904.e. Vector identified by cross-match with the -minscore 20 and -minmatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACAT
BACKWARD: GTTTCCACATCAGCAGC
Plate: 28 row: H column: 24
Seq primer: ATTTAGTGCATATAG.
Location/Qualifiers

FEATURES
source

1..268
/organism:"Sus scrofa"
/db_xref:"taxon:9823"
/clone_lib:"MARC 2P1G"
/tissue_type:"pooled"
/lab_host:"DH10B"
/note:"Vector: PCMV SPOT6; Site 1: XbaI; Site 2: XhoI;
library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 52 a 95 c 80 g 41 t
ORIGIN

alignment_scores:
Quality: 92.50 Length: 81
Ratio: 1.968 Gaps: 4
Percent Similarity: 58.025 Percent Identity: 32.099

alignment_block:
US-09-471-276-831 x AM359511 ..

Align seg 1/1 to: AM359511 from: 1 to: 268

```

16 GlyProValThrGluAlaAlaIlePheTyTGluThrGln***SerLeuTr 32
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
33 GCGCCACTCCCGCCACCCCTCAGTCTAC.....TGACAGGAGAGAGCTG 76
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
32 pAlaGluSerGluHis**LeuLysThrLeuGluGlnCysAspAlaAspV 49
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
77 GGCCTTCTGTGA.....GGTCTGAGCGCCGACGAGAACCTCTCTC 117
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
49 aLProGluPro.....ProGlyAspSerArgLeuProAlaVal 61
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
118 TCCCGACCGGAGACATCTTGAAGGCCAGACGCGG.....GAGGTG 161
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
62 GlnGluTrpGluAlaGlnGluProValHisLeuAspSerProAlaIle 78
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
162 AGGAGCTGGGAGTCCCTCCCGCCATGACTGTGACAGGACCTGTCTTAC 211
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
78 SHISGlnPheLeuLeuThrGluAspThrGlnGlyArgTyArg 92
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
212 ACAGCGCCCTCTCATGTGAGTGGGAGCTCAGATCCAGGAGAGAGA 254
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

seq_name: gb_est1:AM436421

seq_documentation_block:
LOCUS AM436421 273 bp mRNA EST 09-JUL-2000
DEFINITION 76621 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AM436421
VERSION AM436421.1 GI:6971727
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 273)
AUTHORS Fahnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W., and Keele,J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine

JOURNAL COMMENT Unpublished (2000)

CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred

Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCAGCTACGAGC
Plate: 31 row: 0 column: 2
Seq primer: ATTTAGTGACATATAG.
Location/Qualifiers

FEATURES

source

1..273
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2Pig"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 53 a 95 c 82 g 43 t
ORIGIN

alignment_scores:
Quality: 92.50 Length: 81
Ratio: 1.968 Gaps: 4
Percent Similarity: 58.025 Percent Identity: 32.099

alignment_block:

US-09-471-276-831 x AM436421 ..

Align seg 1/1 to: AM436421 from: 1 to: 273

16 G1yProValThrGluAlaAlaIlePheTyGluThrGln**SerLeuTr 32
||||| : : : : : ||| : : : ||
36 GGCCAGTCCCGCCAGCTCAGCTAC.....TGACAGGAGGAGGCTG 79
32 palagIuserGlnHis**LeuIysThrLeuGlyGlnCysAspAlaAspV 49
||||| : : : : : ||| : : : ||
80 GGCGCTTCGTGAA.....GCTGCTGAGCGCCAGGAGAGGCTCTC 120
49 aProGlyPro.....ProGlyAspSerArgLeuProAlaVal 61
: : : : : ||| : : : : : |||
121 TCCCGAGCCCGGAGACTACTTGAAGGCGCAGAGCCG.....GAGGTG 164
62 G1nG1uTrpG1yAlaG1nG1uProValH1sLeuAspSerProAlaIle 78
: : : : : ||| : : : : : |||
165 AGGAGCTGGGGATCCCTCCCGCATGCTGTGAAGACCTCGTTTAC 214
78 sH1sG1nPhelLeuLeuThrG1yAspThrG1nG1yArgTyArg 92
||| : : : : : ||| : : : : : |||
215 ACAGGCCCTCTCATGATGGGAGCTCAGATCCAGGAGCAGA 257

seq_name: gb_est2:B1337778

seq_documentation_block:

LOCUS B1337778 514 bp mRNA EST 30-JUL-2001
DEFINITION 361317 MARC 1Pig Sus scrofa cdna 5', mRNA sequence.
ACCESSION B1337778
VERSION B1337778.1 GI:15031061
KEYWORDS EST.
SOURCE
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 514)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.,
and Keele,J.W.

TITLE Design and use of two pooled tissue normalized cdna libraries for
EST discovery in swine

JOURNAL Unpublished (2000)
CONTACT Smith JPL

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and all trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCAGCTACGAGC
Plate: 127 row: E column: 2
Seq primer: ATTTAGTGACATATAG.
Location/Qualifiers

FEATURES

source

1..514

/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1Pig"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

BASE COUNT 112 a 179 c 125 g 98 t
ORIGIN

alignment_scores:
Quality: 92.50 Length: 81
Ratio: 1.968 Gaps: 4
Percent Similarity: 58.025 Percent Identity: 32.099

alignment_block:

US-09-471-276-831 x B1337778 ..

Align seg 1/1 to: B1337778 from: 1 to: 514

16 G1yProValThrGluAlaAlaIlePheTyGluThrGln**SerLeuTr 32
||||| : : : : : ||| : : : ||
137 GGCCAGTCCCGCCAGCTCAGCTAC.....TGACAGGAGGAGGCTG 180
32 palagIuserGlnHis**LeuIysThrLeuGlyGlnCysAspAlaAspV 49
||||| : : : : : ||| : : : ||
181 GGCGCTTCGTGAA.....GCTGCTGAGCGCCAGGAGAGGCTCTC 221
49 aProGlyPro.....ProGlyAspSerArgLeuProAlaVal 61
: : : : : ||| : : : : : |||
222 TCCCGAGCCCGGAGACTACTTGAAGGCGCAGAGCCG.....GAGGTG 265
62 G1nG1uTrpG1yAlaG1nG1uProValH1sLeuAspSerProAlaIle 78
: : : : : ||| : : : : : |||
266 AGAGCTGGGGATCCCTCCCGCATGCTGTGAAGACCTCGTTTAC 315
78 sH1sG1nPhelLeuLeuThrG1yAspThrG1nG1yArgTyArg 92
||| : : : : : ||| : : : : : |||
316 ACAGGCCCTCTCATGATGGGAGCTCAGATCCAGGAGCAGA 358

seq_name: gb_est1:BE012583

seq_documentation_block:

LOCUS BE012583 544 bp mRNA EST 09-JUL-2000
DEFINITION 122203 MARC 1Pig Sus scrofa cdna 5', mRNA sequence.
ACCESSION BE012583
VERSION BE012583.1 GI:8273508
KEYWORDS EST.
SOURCE
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 544)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,

Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W., and Keele, J.W.
Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail@marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCAGTCACGACG
Plate: 48 row: D column: 21
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers

FEATURES
source 1..548
/organism:"Sus scrofa"
/db_xref:"taxon:9823"
/clone_lib:"MARC 1P1G"
/tissue_type:"pooled"
/lab_host:"DH10B"
/note:"Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI; library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

BASE COUNT 119 a 187 c 144 g 94 t
ORIGIN

alignment_scores:
Quality: 92.50 Length: 81
Ratio: 1.968 Gaps: 4
Percent Similarity: 58.025 Percent Identity: 32.099

alignment_block:

US-09-471-276-831 x BE012583 ..

Align seg 1/1 to: BE012583 from: 1 to: 544

```

16 G1yProValThrGluAlaAlaIlePheTyrGluThrGln**SerLeuTr 32
|||||
104 GGCCCACTCCCGCCACCCCTCAGTCTAC.....TGACAGGAGAGAGCTG 147
32 PAlaGluSerGluHis**LeuLysThrLeuGlyGlnCysAspAlaIaSpV 49
|||||
148 GGCCCTCTCTGAA.....GCTGCTGAGGCCACAGGAGAGAGCTCTC 188
49 a1ProGlyPro.....ProGlyAspSerArgLeuProAlaVal 61
|||||
189 TCCCCACCCCGGAGACACTTCAAGGAGCCAGACCCG.....GAGGTG 232
62 GlnGluTrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleLys 78
|||||
233 AGGAGCTGGGAGATCCCTCCCATCAGCTGAGAGACCTCGTTTAC 282
78 sHisGlnPheLeuLeuThrGlyAspThrGlnGlyArgTyrArg 92
|||||
283 ACACGCCCTCTCATGATGGGAGACTCAAGATCCAGGAGACAGA 325

```

seq_name: gb_est2:BG733169

seq_documentation_block: 548 bp mRNA EST 11-MAY-2001
LOCUS BG733169
DEFINITION 346903 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BG733169
VERSION BG733169.1 GI:14019453
KEYWORDS EST.
SOURCE p1g.
ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 548)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W., and Keele, J.W.
Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail@marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCAGTCACGACG
Plate: 108 row: H column: 14
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers

FEATURES
source 1..548
/organism:"Sus scrofa"
/db_xref:"taxon:9823"
/clone_lib:"MARC 1P1G"
/tissue_type:"pooled"
/lab_host:"DH10B"
/note:"Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI; library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

BASE COUNT 119 a 187 c 149 g 93 t
ORIGIN

alignment_scores:
Quality: 92.50 Length: 81
Ratio: 1.968 Gaps: 4
Percent Similarity: 58.025 Percent Identity: 32.099

alignment_block:

US-09-471-276-831 x BG733169 ..

Align seg 1/1 to: BG733169 from: 1 to: 548

```

16 G1yProValThrGluAlaAlaIlePheTyrGluThrGln**SerLeuTr 32
|||||
87 GGCCCACTCCCGCCACCCCTCAGTCTAC.....TGACAGGAGAGAGCTG 130
32 PAlaGluSerGluHis**LeuLysThrLeuGlyGlnCysAspAlaIaSpV 49
|||||
131 GGCGCTCTCTGAA.....GCTGCTGAGGCCACAGGAGAGAGCTCTC 171
49 a1ProGlyPro.....ProGlyAspSerArgLeuProAlaVal 61
|||||
172 TCCCCACCCCGGAGACACTTCAAGGAGCCAGACCCG.....GAGGTG 215
62 GlnGluTrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleLys 78
|||||
216 AGGAGCTGGGAGATCCCTCCCATCAGCTGAGAGACCTCGTTTAC 265
78 sHisGlnPheLeuLeuThrGlyAspThrGlnGlyArgTyrArg 92
|||||
266 ACACGCCCTCTCATGATGGGAGACTCAAGATCCAGGAGACAGA 308

```

seq_name: gb_est2:T68210

seq_documentation_block: 97 bp mRNA EST 22-FEB-1995
LOCUS T68210
DEFINITION yc40b03.r1 Stratiogene liver (#937224) Homo sapiens cDNA clone IMAGE:83117 5' similar to SP:A1BG_HUMAN P04217 ;, mRNA sequence.

```

ACCESSION      T68210
VERSION        T68210.1  GI:6793358
KEYWORDS
SOURCE
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS        Hillier,L., Lennon,G., Becker,M., Bonaudo,M.F., Chiappelli,B.,
                Chissee,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins
                ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
                ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
                Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Treviski,E.,
                Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
                Generation and analysis of 280,000 human expressed sequence tags
                Genome Res. 6 (9), 807-828 (1996)
                97044478
TITLE          Contact: Wilson RK
JOURNAL        Washington University School of Medicine
MEDLINE        4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
COMMENT        Tel: 314 286 1800
                Fax: 314 286 1810
                Email: est@watson.wustl.edu
                Insert Size: 87
                High quality sequence starts: 1 High quality sequence stops: 1
                Source: IMAGE Consortium, LBNL This clone is available royalty-free
                through LBNL ; contact the IMAGE Consortium (info@image.lbnl.gov)
                for further information. Trace considered overall poor quality
                Possible reversed clone: similarity on wrong strand
                Insert length: 87 Std Error: 0.00
                Seq primer: M13RPI
                High quality sequence stop: 1.
FEATURES
SOURCE
                1..97
                location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="GDB:500174"
                /db_xref="taxon:9606"
                /clone IMAGE:83117"
                /clone_id="Stratagene liver (#937224)"
                /sex="male"
                /dev_stage="49 years old"
                /lab_host="SOLR cells (kanamycin resistant)"
                /note="Organ: liver; Vector: pBluescript SK. Site:1: EcoRI
                ; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo
                dt. Hepatectomy from normal male caucasian. Average insert
                size: 1.1 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5'
                GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
                CTCGACGTTTTTTTTTTTTTTT 3'"
BASE COUNT     24 a 26 c 23 g 19 t 5 others
ORIGIN
alignment_scores:
                Quality: 92.00 Length: 32
                Ratio: 3.538 Gaps: 0
                Percent Similarity: 81.250 Percent Identity: 75.000
alignment_block:
US-09-471-276-831 x T68210 ..
Align seg 1/1 to: T68210 from: 1 to: 97
23 TlePhetYrCluThrGln**SerLeuTrpAlaGluSerGluHis***Le 39
|||||
2 ATATTTTATGAGCGACGCCACCTCTGTAGAGAGAGATCCNAATCACTGCT 51
39 ulYsthrLeuGlyGlnCysAspAlaAspAlaProGlyProProGly 54
1 |||||
52 NAA.ACCCTTAGCCCAATGTGACGNTGACGNCAGCAGCCACCTGGA 96
seq_name: gb_est2:BG819104
seq_documentation_block:

```

```

LOCUS       BG819104       809 bp       mRNA
DEFINITION  B627611193f1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4932190
ACCESSION   BG819104
VERSION     BG819104.1  GI:14166691
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 809)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1995)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            CDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLNL0857 row: 1 column: 23
            High quality sequence stop: 800.
FEATURES
     source
         1..809
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4932190"
            /clone_lib="NCI_CGAP_Brn67"
            /tissue_type="anaplastic oligodendroglioma with 1p/19q
            loss"
            /lab_host="DH10B (T1 phage-resistant)"
            /note="Organ: brain; Vector: pCMV-SPOrt6; Site1: NOT1;
            Site2: Salt; Cloned unidirectionally. Primer: Oligo d
            Average insert size 2.3 kb. Constructed by Life
            Technologies. Note: This is a NCI_CGAP Library."
BASE COUNT   252 a      180 c      180 g      197 t
ORIGIN
alignment_scores:
    Quality:      91.00      Length:      75
    Ratio:        2.022      Gaps:        2
Percent Similarity: 60.000      Percent Identity: 34.667
alignment_block:
US-09-4771-276-831 x BG819104/rev ..
Align seg 1/1 to reverse of: BG819104 from: 1 to: 809
50 ProGlyProProGlyAspSerArgLeuProAlaValGlnGlnTrpGlyAl 66
|||||
|||||
231 CGAGCGACAGAGTCGTATAGGAGGCTTCCTCCCTGGCCCTGGGGA.. 184
|||||
66 aGlnGlnProValHisLeuAspSerProAlaIleuHisGlnPheLeuL 83
|||||
163 ...CATGTGATACAGGTGACGACCTCCATCTTTAAATGAGAGTTTAACT 138
|||||
83 euThrGlyAspThrGlnGlyArgTyrArgGlySargSerGlyLeuSerThr 99
|||||
137 TGACTGCCAGAAACAGCATTTCTTAT.....TTCTTCCACAGC 100
|||||
100 GlyTrp**GlnLeuSerLysLeuLeuGlnLeuThrGlyProLysValLe 116
|||||
99 TCAGTGCACACTTCAGCTGTCCTGACAGCTCCCTTCCTCCAAATCTCT 50
|||||
116 uAlaCysSerLeuAlaLeuAspGly 124
|||||
49 AAACGTTGCTGCTGCCCGCAGATGCA 25
seq_name: gb_est2:BG75650

```



```

17  ovalThrGluAlaAlaIlePheThrGln**SerLeuTrpAlaG 34
|||||: ||| :|||: |||: |||
1002 AGTGTGATGATGCTCTGTCGACGCCAATGTTGCTACCGCAGTTGG... 957
34  luserGluHis**LeuLysThrLeuGlyGln..... 44
||||| :|||: |||: |||: |||
956 .....CGACACGTGGTGTGAGTCATCTCTGCTGTAGATTGGGTAGTG 912
45  ...CysAspAlaAspValProGlyProGly...ArgLeuProAl 60
||||| ||| ||||| ||| ||| :|||
911 GTCTCTGCTGTGATAGTCGCGGAGCCCGCGGCTCATGTCGCGCAC 862
60  aValGlnGluTrpGlyAlaGlnGluProValHis.....LeuAspSerP 75
|||||: ||||| :|||: |||
861 GGTATCCCTTGGGGGCTTCCCAACCTCTGCTTCCACCCGTGTGAGATC 812
75  roAlaIleLysHisGlnPheLeuLeuThrGlyAspThrGln..... 88
||| ||| |||||: |||: |||
811 CTCATATTGG...CAGTTTCTGTGAGCGGCGAGTCGAGGCCACGACG 765
89  .....GlyArgTyrArgCysArgSerGlyLeuSerThrG1 100
||||| ||| ||||| |||||
764 AACACGTGGCAAGTGGGACCGGACGCAATCCAGTGG..... 726
100 YTrp**GlnLeuSerLysLeuLeuGluLeuThrGlyProLysValLeu 116
||| :||| :|||: |||: |||
725 .....CTTCCCGCTCAGCTCAGAGACTGCGCCACACACCTTGG 687
seq_name: gb_gss:A0400188

seq_documentation_block:
LOCUS A0400188 500 bp DNA GSS 06-MAR-1999
DEFINITION mgx0006L22f CUG1 Rice Blast BAC library Magnaporthe grisea genomic
clone mgx0006L22f, DNA sequence.
ACCESSION A0400188
VERSION A0400188
KEYWORDS GI:4371215
SOURCE GSS.
ORGANISM Magnaporthe grisea.
Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes Incertae sedis; Magnaportheaceae; Magnaporthe.
1 (bases 1 to 500)
Yu.Y., Zhu.H., Boyd.C.A., Gaudette.B., Gayle.A., Kingsbury.R.,
Phillips.K., Sasnowski.M., Wing.R.A. and Dean.R.A.
A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Dean RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson University, Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdeane@clemson.edu
Seq primer: TAATACGACTCACTATAGG
Class: BAC ends
High quality sequence stop: 367.
Location/Qualifiers
1..500
/organism="Magnaporthe grisea"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgx0006L22f"
/clone_lib="CUG1 Rice Blast BAC library"
/issue_type="Protoplasts"
/lab_host="E. coli DH10B"
/note="Vector: pBACW10H; Site_1: HindIII; Site_2: HindIII;
Rice blast is one of the most devastating fungal diseases
of rice world wide. It is a filamentous ascomycete with
a haploid genome (n=7) of approximately 40 Mbp. Rice
blast is an important model fungal pathogen for studying
numerous aspects of the fungal-host interaction. In

```

order to facilitate genome wide analysis, a BAC library containing 9216 clones with an average insert size of 130 kbp was constructed. This library represents greater than 25x genome coverage. High density colony filters are available upon request.

BASE COUNT 147 a 131 c 98 g 124 t
ORIGIN

alignment_scores:
Quality: 89.00 Length: 114
Ratio: 1.534 Gaps: 5
Percent Similarity: 50.877 Percent Identity: 33.333

alignment_block:
US-09-471-276-831 x A0400188/rev ..

Align seq 1/1 to reverse of: A0400188 from: 1 to: 500

```

6  ValPheLeuLeuTrpGlyValThrTrpGlyProValThrGluAla1 22
   ::|||: ||| :|||: |||: |||
388 CTCTTTTCTATCTCTTGCGATATCTTGGGCTCTGCTCGCCGACGC 339
22  aIlePheTyrGluThrGln**SerLeuTrpAlaGluSerGluHis**L 39
   |||: ||||| |||||: |||: |||
338 TTTTCTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 289
39  euLys.....ThrLeuGlyGlnCysAspAlaAspValProGlyPro 53
   |||: ||||| ||||| :|||
288 TTATGTGCGCAACGTGTGCTGCGGAGTTAANAAGTA..... 251
54  GlyAspSerArgLeuProAlaValGlnGluTrpGlyAlaGlnGluPro 70
   ||| ||| ||||| ||||| :|||: |||
250 .....TGATATTACTATGATGTCAGAAATGCAAAATGCAAAACCG 207
70  lHisLeuAspSerProAlaIleLysHisGlnPheLeuLeuThrGlyAs 87
   |||: ||||| ||||| :|||: |||
206 CCGTTTGATGCCGATCCCATCAAGGCT.....CTGTGTCAGAGCAC 163
87  hrgInGlyArg.....TyrArgCysArgSerGlyLeuSerThrGly 100
   ||||| :||| :|||: |||: |||
162 GTATGCGAAGATCGATCGCGGCTGTGAGTCTGGGATG..... 122
101 Trp**GlnLeuSerLysLeuLeuGluLeuThrGlyProLys 114
   |||||: ||||| :|||: |||
121 .....TTGTAGATCTGTTGGACCAAAA 98
seq_name: gb_est:AW056411

seq_documentation_block:
LOCUS AW056411 233 bp mRNA EST 01-DEC-1999
DEFINITION priu-4Q Representational difference Analysis Products Rattus
norvegicus cDNA similar to murine cDNA accn AA530090, mRNA
sequence.
ACCESSION AW056411
VERSION AW056411
KEYWORDS GI:6501051
SOURCE EST.
ORGANISM Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 233)
Tollet-Egnell,P., Flores-Morales,A., Odeberg,J., Lundberg,J. and
Norstedt,G.
Differential cloning of growth hormone regulated hepatic
transcripts in the aged rat
JOURNAL Unpublished (1999)
COMMENT Contact: Tollet-Egnell, P.
Department of Molecular Medicine
Karolinska Institute
CMM, 18:01, Karolinska Hospital, 171 76 Stockholm, Sweden
Tel: 46 8 51774395

```



```

seq_documentation_block:
LOCUS      BG178413      892 bp      mRNA      EST      06-FEB-2001
DEFINITION BG230201F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4431264 5',
            mRNA sequence.
ACCESSION  BG178413
VERSION    BG178413.1 GI:12685116
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE  1 (bases 1 to 892)
            NIH-MGC http://mgs.ccl.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DRP
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM10186 row: b column: 01
            High quality sequence stop: 623.
FEATURES
    source
        1..892
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_image="4431264"
            /clone_lib="NIH_MGC_91"
            /tissue_type="adenoarcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.4 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC library."
BASE COUNT      199 a      275 c      247 g      171 t
ORIGIN
alignment_scores:
    Quality:      87.50      Length:      110
    Ratio:        1.509      Gaps:      5
    Percent Similarity: 52.727      Percent Identity: 35.455
alignment_block:
US-09-471-276-831 x BG178413 ..
Align seg 1/1 to: BG178413 from: 1 to: 892
12 G1yAlhTrTPgIyProValThrGluAlaAlaIlePheTyrGluThrG1 28
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 GGCGTGGAGTGGCGCCAGACCCAGTGCAGGACGAGGAGCCGCCCA 247
28 n***SerLeuTrpAlaGluSerGluHis***LeuLysThrLeuGly Gln 44
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
248 GCCCACACTCTGGCTGAGCGGCTGTGCATC...ACCGAGGAGATGCC 294
45 CysAspAlaAspValProGlyProGlyIleAspSerArgLeuProAlaIle 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 CGTACACCTCTGGTGTTCAGGGGATCTGGAGACCCAGAGATGCCGTGT 344
61 L.Ginglu.....TrpGlyAlaGingluProValHisLeu 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 ATAGAGAAAGAAAGAACAGACACCTGGATACACGATGCCACAGAGATT 394
73 AspSerProAlaIleLysHisGlnPhe.....LeuLeuThrGlyAs 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 .....GTGAAGAAAGGCCAGTTCGCCATCCCGTCACATCACCCTGGA 435

```

```

86 pThnGInG1ArgTyrArgCysArgSerGlyLeuSerThr...GlyTrp* 102
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
436 ACACACAGGGCGGATGCTGCTTTCTACGCTACACACTGCAGGCTGT 485
102 *GlnLeuSerLysLeuLeuGlnLeu 110
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
486 CAGAGCCCACTGAGACCCCTGGAGCTG 511
seq_name: gb_est1:AL532540
seq_documentation_block:
LOCUS      AL532540      920 bp      mRNA      EST      13-FEB-2001
DEFINITION AL532540 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DW006Y106 5
            prime, mRNA sequence.
ACCESSION  AL532540
VERSION    AL532540.1 GI:12796033
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE  1 (bases 1 to 920)
            Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
            Full-length cDNA libraries and normalization
            Unpublished (2001)
JOURNAL    Contact: Genoscope
COMMENT    Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
    source
        1..920
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_image="CS0DW006Y106"
            /clone_lib="LTI_NFL001_NBC4"
            /sex="male"
            /tissue_type="neuroblastoma cells"
            /lab_host="DH10B"
            /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed
            by Life Technologies. Contact : Feng Liang Life
            Technologies, a division of Invitrogen 9800 Medical Center
            Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
            8371 Email : filang@lifetech.com URL :
            http://fulllength.invitrogen.com"
BASE COUNT      199 a      297 c      261 g      156 t      7 others
ORIGIN
alignment_scores:
    Quality:      87.50      Length:      115
    Ratio:        1.620      Gaps:      5
    Percent Similarity: 46.957      Percent Identity: 30.435
alignment_block:
US-09-471-276-831 x AL532540 ..
Align seg 1/1 to: AL532540 from: 1 to: 920
12 G1yAlhTrTPgIyProValThrGluAlaAlaIlePheTyrGluThrG1 28
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
208 GGCGTGGAGTGGCGCCAGACCCGATGCATGCAGGAGGCCCTTCCCAA 257
28 n***SerLeuTrpAlaGlu.....SerGluH 37
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
258 ACCACACCTCTGGGCTGAGCAGGCTGTGTATATAGCTGGGGAGCCCG 307
37 is***LeuLysThrLeuGlyGlnCysAspAla.....Asp 48
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
308 TGACCATCTGTGTGTAGGGAGAGCTGGAGGCCAGAGATACCACTGGAT 357

```


Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DRP/Gazdar
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1CM1237 row: a column: 17
 High quality sequence stop: 666.
 Location/Qualifiers

FEATURES

source

```
1..874
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4547992"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: Lung; Vector: pOT7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAGC(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

BASE COUNT 210 a 200 c 274 g 190 t
 ORIGIN

alignment_scores:

Quality: 86.00 Length: 49
 Ratio: 2.966 Gaps: 2
 Percent Similarity: 59.184 Percent Identity: 36.735

alignment_block:

US-09-471-276-831 x BG330746/rev ..

Align seg 1/1 to reverse of: BG330746 from: 1 to: 874

```
49 ValProGlyProProGlyAspSerArgLeuProAlaValGlnGluTrpG1 65
|||||
760 GTCCCTGGCCCCCTCTGACCCCAAGCTGCTGCTGCTGCTGCCATGCGCC 711
|||||
65 yAlaGlnGluProValHisLeuAspSerProAlaIleLysHisGlnPheL 82
|||||
710 CTGGCCCCAGCTCCCACTGACACCACTCTCTCTCCACGCG.... 665
|||||
82 euleuthrGlyAspThrGlnGlyArg.....TyrArgCys 93
|||||
664 .....ACCCAGGGCCAGCGCTGCGATTTCTCCATCGGTGC 632
```

seq_name: gb_est2:BG365440

seq_documentation_block: 797 bp mRNA EST 08-MAR-2001
 LOCUS BG365440
 DEFINITION HVSM10002107f Hordeum vulgare 20 DAP spike EST library HVCDA0010
 (20 DAP) Hordeum vulgare cDNA clone HVSM10002107f, mRNA sequence.
 ACCESSION BG365440
 VERSION BG365440.1 GI:13254539
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Hordeum.
 1 (bases 1 to 797)
 Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
 ,Y., Anderson,H., Dale,J., Henry,D., Kennodle,S., Palmer,M., Rambo
 ,T., Sakai,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
 Wood,T.
 Development of a genetically and physically anchored EST resource
 for barley genomics

JOURNAL Unpublished (2000)
 COMMENT Contact: Ming RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu

Seq primer: AATTACCCCTCACTAAGG
 High quality sequence stop: 390.
 Location/Qualifiers

FEATURES

source

```
1..797
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSM10002107f"
/clone_lib="Hordeum vulgare 20 DAP spike EST library"
/HVCDA0010 (20 DAP)"
/tissue_type="20 DAP spike"
/lab_host="SOLR"
/notes="Vector: lambdaZAP; Site:1: EcoRI; Site:2: XhoI; For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/ To
order a clone see http://www.genome.clemson.edu/orders"
```

BASE COUNT 134 a 194 c 277 g 192 t
 ORIGIN

alignment_scores:

Quality: 85.50 Length: 82
 Ratio: 1.943 Gaps: 5
 Percent Similarity: 53.659 Percent Identity: 39.024

alignment_block:

US-09-471-276-831 x BG365440 ..

Align seg 1/1 to: BG365440 from: 1 to: 797

```
48 AspValProGlyProProGlyAspSerArgLeuProAlaValGlnGluTr 64
|||||
305 GAGCGCCGAGCCCTCCGAGAGATGCGCTGAGACCTCTACGG.... 349
|||||
64 pGlyAlaGlnGluProValHisLeuAspSerProAlaIleLysHisGlnP 81
|||||
350 ....GCCACAGCTCCCTCAACGAGAT.....CAAT 377
|||||
81 heLeuLeuthrGlyAspThrGlnGlyArgTyrArgCysArgSerGlyLeu 97
|||||
378 TCCGACGTTCCGGC.....TCCGCGCGCGCTTCCAGACGAGTCTAC 418
|||||
98 SerThrGlyTyr***.....GlnLeuSerLysLeuLeu.. 108
|||||
419 GGCACACCTGCGACGTGAGGAGCCATGCTTGGATGTCGCGCTTACTCC 468
|||||
109 .....GlnLeuthrGlyProLysValLeuAlaCys 118
|||||
469 ATCTTGGCGTGGTGTGCGCATGATGATGCTGCTTCCGCGTGC 514
```

seq_name: gb_est2:BG337700

seq_documentation_block: 945 bp mRNA EST 27-FEB-2001
 LOCUS BG337700
 DEFINITION 602435261f1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:455199 5',
 mRNA sequence.
 ACCESSION BG337700
 VERSION BG337700.1 GI:13144138
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 945)
 NIH-MGC <http://mgs.ncl.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1250 row: 1 column: 16
High quality sequence stop: 726.
Location/Qualifiers

FEATURES
source 1..945

/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/clone:"IMAGE:4553199"
/clone_11b:"NIH_MGC_46"
/tissue_type:"leiomyosarcoma cell line"
/lab_host:"DH10B (phage-resistant)"
/note:"Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH_MGC library."

BASE COUNT 216 a 219 c 285 g 225 t

ORIGIN

alignment_scores: Quality: 85.00 Length: 89
Ratio: 1.771 Gaps: 3
Percent Similarity: 53.933 Percent Identity: 31.461

alignment_block:

US-09-471-276-831 x BG337700/rev ..

Align seg 1/1 to reverse of: BG337700 from: 1 to: 945

```

41  ThcLeuGlcInGlyAspAlaAspValProGlyProGlyAspSerAr 57
|||||  ::  :::::  |||||  ||:::||||
774  ACACTGCACGCTCCACACGACGACGACGACGACGACGACGACG 725
57  gLeuProAlaValGlnGluTrpGlyAlaGlnGluPro.ValHis..... 71
:::||||:  ::  :::::  |||  |||
724  TGTACCCGACATGACGATGTCCTTGGCGTCCGCCACACGATGATAC 675
|||||  :::  :::::  |||||
72  LeuAspSerProAlaLeuHisLeuGlnPheLeuLeuThrGlyAspTrpG1 88
|||||  :::  :::::  |||||
674  CAACTTCTCCACGACGACGACGACGACGACGACGACGACGACG 643
88  nGlyArgTrpArgGlyAspSerGlyLeuSerThrGlyTrp**GlnLeu 105
|||||  :::  :::::  |||||  |||  :::||||
642  ACCCGCGCTTTCTTCACGCTGCTGCTTCGCAATCTTGGCTGATCTTT 593
105  eGlyLeuLeuGlnGluLeuThrGlyProGlyValLeuAlaGlySerLeuAla 121
|||||  :::  :::::  |||||  :::  :::
592  CTGAACATCATCACATCACAGAGATCAAAATTCATATGCTTGCAAAAGC 543
122  LeuAspGlyAlaSer 126
|||  |||||  :::
542  CTTAAAGGTGCAAT 528

```

seq_name: gb_est2:BG396925

seq_documentation_block: 886 bp mRNA EST 12-MAR-2001
LOCUS BG396925
DEFINITION 602433829P1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4552014 5',

ACCESSION mRNA sequence.
VERSION BG396925
KEYWORDS GI:13290373
SOURCE Est.
ORGANISM human.

REFERENCE 1 (bases 1 to 886)
NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DFP

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1247 row: 1 column: 07
High quality sequence stop: 702.
Location/Qualifiers

FEATURES
source 1..886

/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/clone:"IMAGE:4552014"
/clone_11b:"NIH_MGC_20"
/tissue_type:"melanotic melanoma"
/lab_host:"DH10B (phage-resistant)"
/note:"Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 215 a 203 c 263 g 205 t

ORIGIN

alignment_scores: Quality: 84.50 Length: 80
Ratio: 1.837 Gaps: 4
Percent Similarity: 57.500 Percent Identity: 32.500

alignment_block:

US-09-471-276-831 x BG396925/rev ..

Align seg 1/1 to reverse of: BG396925 from: 1 to: 886

```

50  ProGlyProProGlyAspSerArg...LeuProAlaValGlnGluTrpG1 65
|||||  :::  :::::  |||||  |||  :::||||
746  CCAAGCGCTCCACAGCCTCAGCAAGTGTACCCACACGACGATCTGGCG 697
65  yAla...GlnGluProValHisLeuAspSerProAlaHisHisHisLnp 81
|||||  :::  :::::  |||||  :::  :::||||
696  TGCCTCCGCCACATGATGATGATGATGATGATGATGATGATGATGAT 648
81  heLeuLeuThrGlyAspTrpGlnGlyArgTrpArgGlySerGly... 96
|||||  :::  :::::  |||||  |||  :::||||
647  ..... ACCCAGACCCGCTTTCTTACCGCTGCTGTTTG 615
97  LeuSerThrGlyTrp**GlnLeuSerIysLeuLeuGluLeuThrGlyPr 113
|||  |||  :::  :::::  |||||  :::  :::
614  CTTCGCAATCTTGTGGCTTGTGATCTTGTGCAATCATCAATCATCAAGCA 565
113  oLysValLeuAlaGlySerLeuAlaLeuAspGlyAlaSer 126
:::  :::  :::  |||||  :::
564  TCAAAATTCATATGCTTGCAAAAGCCTTAAAGTGCCAAAT 525

```

```

seq_name: gb_est1:BE542780
seq_documentation_block: 1025 bp mRNA EST 09-AUG-2000
LOCUS BE542780 mRNA EST 09-AUG-2000
DEFINITION 601066385f1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3452661 5',
mRNA sequence.
ACCESSION BE542780
VERSION BE542780.1 GI:9771425
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1025)
TITLE NIH-MGC http://www.ncbi.nlm.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHM8434 row: 3 column: 22
High quality sequence stop: 150.
Location/Qualifiers
1..1025
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3452661"
/clone_lib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
294 g 217 t

BASE COUNT 295 a 219 c 217 t
ORIGIN
1025

alignment_scores:
Quality: 84.00 Length: 98
Ratio: 1.750 Gaps: 4
Percent Similarity: 48.980 Percent Identity: 28.571

alignment_block:
US-09-471-276-831 x BE542780 ..

Align seg 1/1 to: BE542780 from: 1 to: 1025
10 LeuTrpGlyValThrTrpGlyProValThrGluAlaAlaIlePheTyrG1 26
||||| ||||| :|||: ||||| ||
457 CTGCGCATTCGATTCGGGGCGGATCTTCAAAAAGCTCATATTCACAGA 506
26 uTrGln***SerLeuTrpAlaGluSerGluHis***LeuLysThrLeuc 43
| ||||| ||||| :|||: |||||
507 ACGGGCTCTCTCTTGGGCGACAA.....ATATTGAACACCTA. 546
43 LysGlnGSAAPAlaAspValProGlyProPro..... 53
547 ..ACGTGTGGGCGGTGTGCTCCGCCACAGTGTCTGGGGGTATTGT 594
54 .....GlyAspSerArgLeuPr 59
595 TTTCACAAAGCCCTGTGGGGGTCTTCCACACACTTTACAAAGAGAAC 644
59 cAlaValGlnGluTrpGlyAlaGlnGluProValHisLeu.....A 73
||||| :|||: ||||| |||
645 TACTGTATAAGTTCCGGGGGAGAGTCCCTCCACGCTGGGAGGGA 694

```

```

73 spSerProAlaIleLysHisGlnPheLeuThrGlyAspThr 87
||||| :|||: ||||| |||
695 ACACACCCCTCTGTCTACAGATATAGGGGACAGCGCGGACA 738
seq_name: gb_est1:AM001162
seq_documentation_block: 481 bp mRNA EST 27-OCT-1999
LOCUS AM001162 mRNA EST 27-OCT-1999
DEFINITION w25c03.x1 Soares_Dieckgraefe_colon_NHCD Homo sapiens cDNA clone
IMAGE:2521060 3', mRNA sequence.
ACCESSION AM001162
VERSION AM001162.1 GI:5848078
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 481)
TITLE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Glibco
High quality sequence stop: 462.
Location/Qualifiers
1..481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2521060"
/clone_lib="Soares_Dieckgraefe_colon_NHCD"
/tissue_type="colonic mucosa from 3 patients with Crohn's
disease"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pTrp3d-Pac (Pharmacia) with a
modified polylinker; Site: 1; Not I; Site: 2; Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTTACCAATCTGAGTGGAGCGCGCCGCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTrp3 vector. Library
went through one round of normalization. Tissue samples
provided by Dr. Brian Dieckgraefe (Washington University,
dieck@im.wustl.edu); colonic mucosa represents a range of
disease involvement from moderate to severe Crohn's
disease; samples include both perforating (fistulas) and
non-perforating samples. Library constructed by Bento
Soares and M. Fatima Bonaldi."
128 c 127 g 124 t

BASE COUNT 102 a 128 c 127 g 124 t
ORIGIN
1025

alignment_scores:
Quality: 83.50 Length: 112
Ratio: 1.704 Gaps: 5
Percent Similarity: 43.750 Percent Identity: 27.679

alignment_block:
US-09-471-276-831 x AM001162 ..

Align seg 1/1 to: AM001162 from: 1 to: 481
22 AlaIlePheTyrGluThrGln***SerLeuTrpAlaGluSer..... 35
||| ||||| ||||| :|||: |||||
37 GCAAAATTTTACGAAGAAGTCAGATTCACCCGTGAGCAGACACCTTTT 86
36 .....GlnHis 37
87 CTCACGGAGAGACTGTGTACACAGTCCGTGTGGCGGTTCATCAT 136

```



```

DEFINITION      206184 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION       BE753245
VERSION         BE753245.1
KEYWORDS        EST.
SOURCE          Bovidae; Bovinae; Bos.
ORGANISM        Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE       1 (bases 1 to 535)
AUTHORS        Smith,T.P., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
                Casas,E., Wray,J.E., White,J., Cho,J., Fahrentz,S.C., Bennett,
                G.L., Hooton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-Mckown,C.G.,
                Petzel,C., Holt,I., Karamycheva,S., Liang,F., Quakenbush,J. and
                Keele,J.W.
TITLE           Sequence evaluation of four pooled-tissue normalized bovine cDNA
                libraries and construction of a gene index for cattle
JOURNAL         Genome Res. 11 (4), 626-630 (2001)
MEDLINE         21180013
COMMENT         Contact: Smith TPL
                USDA, ARS, US Meat Animal Research Center
                PO Box 166, Clay Center, NE 68933-0166, USA
                Tel: 402 762 4366
                Fax: 402 762 4390
                Email: smith@email.marc.usda.gov
                Single pass sequencing. Bases called and alt-trimmed with phred
                v0.980904.e. Vector identified by cross-match with the -minscore 18
                and -mismatch 12 options.
                PCR Primers
                FORWARD: AGCAACACGCTATGACCAT
                BACKWARD: GTTTTCCAGCTCAGCAGC
                Plate: 47 row: E column: 20
                Seq primer: ATTTAGTGACGACTATAC.
FEATURES        Location/Qualifiers
                source          1..535
                                /organism="Bos taurus"
                                /db_xref="taxon:9913"
                                /clone_lib="MARC 2BOV"
                                /tissue_type="pooled"
                                /lab_host="DH10B"
                                /note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
                                library made from pooled tissue from testis, thymus,
                                semitendinosus muscle, longissimus muscle, pancreas,
                                adrenal, and endometrium."
BASE COUNT      125 a 159 c 133 g 118 t
ORIGIN
alignment_scores:
    Quality:      82.50          Length:      142
    Ratio:         1.352          Gaps:         9
    Percent Similarity: 42.958    Percent Identity: 28.873
alignment_block:
US-09-471-276-831 x BE753245 ..
Align seg 1/1 to: BE753245 from: 1 to: 535
2 SerMetLeuValValPheLeuLeuLeuTP.....GlyValThrTr 15
:::|||||:::|||||:::|||||:::|||||:::|||||
131 ACCATCTCTCGGACGCTTCACTCTTACTCTTTTGTGCTGCACTGCTC 180
15 pcjYpova1Thrg1ua1a1a1ePheThrg1uThrg1n**SerLeuY 32
||| ||||| |||:::|||||:::|||||
181 GGST...GFGACACCAATGCACTGATG...CCATCTCAACCGACTAT 224
32 rPa1a1uSer1uH1u1s**Leu1yThrg1u1y1nc1yasp1a1asp 48
|| |||||:::|||||
225 GCATTGAGTCAACTAC.....ATGACACTCTGG... 241
49 ValProG1YpProProG1YpSerATrgLeuPro1a1a1a1a1nc1uTrpG1 65
||| |||:::|||||
242 ...CCCCAGGCCCTTGGAGAGAC.....ATGACACTCTGG... 274

```

```

65 yAlaGInGluProValHisLeuAspSerProAla... IleYshHisGInp 81
275 .....TGTAAAGCCCTCTCAGATTTTCAAGCAAGT 305
81 heLeuLeuThrGlyAsp.....ThrGln..... 86
|||||||
|||||
306 TCTTGTTCGTCGAAGGATAAGACACAGATGACCTGCAGTCCGCCCTTCCCGA 355
87 ..... 88
356 AAGACCTTCCAGTTTCATTCCTCCATCGTCGCCCTTACTCAGCGCAATATAC 405
89 .GlyATGTTATGCGysArGysrGlyLeuSorThGlyTTrp**GInLeuS 105
|||||
|||||
406 AGGCTTTTACAGGTCGTCTACTGGAAGAGACAGGCTGTAAGGCCA 455
105 eRyLSLeuLeuGluLeuThrGlyPro 113
|||||
|||||
456 GTAAAGCTTTAGACTTGAGGACACCA 481

seq_name: gb_est2:BG442106

seq_documentation_block:
LOCUS BG442106 792 bp mRNA EST 15-MAR-2001
DEFINITION GA_Ea0015M19f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboeum cDNA clone GA_Ea0015M19f, mRNA sequence.
ACCESSION BG442106
VERSION BG442106.1 GI:13351758
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 792)
Wing,R.A., Fritsch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: TAATACGACTCACTATGAGG
High quality sequence stop: 719.
Location/Qualifiers
1..792
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0015M19f"
/clone_1lb="Gossypium arboreum 7-10 dpa fiber library"
/clone_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="Vector: pBR-CMW; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 219 a 142 c 202 g 229 t
ORIGIN

alignment_scores:
Quality: 82.50 Length: 114
Ratio: 1.500 Gaps: 6
Percent Similarity: 48.246 Percent Identity: 32.456

alignment_block:
US-09-471-726-831 x BG442106

```

Align seg 1/1 to: BG442106 from: 1 to: 792

```

18 ValThrcIu...AlaAlaIlePheTyrgIuThrcIn**SerLeuTrpAl 33
   ||||| ||||| ||||| ||||| |||||
195 TTGACAGAAAGGGGCCCTTGAAATTTGGTACACA..... 230
33 agIuSerGIuHis**LeuLysThrLeuGlyGInCysAspAlaAspValP 50
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 .....CATGGATTGCATGGTGGACGGTGATTCCTAGTTGGGTG 270
50 roGIuPro.....ProGIuAspSerArgLeuProAla 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
271 TTGCTCCTTCATATGCCCAATTCCTGGCTCAGCGCCTGTCTCATTC 320
61 ValGIuInIuTrpGIuAlaGInIuPro..... 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
321 GCCTGTGCTCCTAGGACCAATCGAATATAGTCTCCTGTGATGCATC 370
70 ....ValHisLeuAspSerProAlaIleLysHisGInPheLeuThr. 84
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 GATGTGATCATTTGATGACTTGGCAGGCAATATATCTCTGCTGAAT 420
85 .GIuAspThrGInGlyArgTyArgCysArgSerGIuLeuSerThrGIy 100
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
421 ATCCAGATGCAAAAGAAATACATTTG.....TCTTCGCAC 458
101 Trp**GIuLeuSerLysLeuLeuGInLeuThrGIyProLys 114
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
459 ACCATAGCCTTGAGAGATTGTGTGAATTTCTTGTCGCAAG 500
seq_name: gb_est2:BG177264

```

```

seq_documentation_block: 898 bp mRNA EST 06-FEB-2001
LOCUS BG177264 60231452F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4420261 5',
DEFINITION mRNA sequence.
ACCESSION BG177264
VERSION BG177264.1 GI:12683967
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia: Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 898)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgsrbs@mail.nih.gov
Tissue Procurement: Louis Staudt, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Place: LHAM10157 row: g column: 14
High quality sequence stop: 703.
Location/Qualifiers
1..898
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4420261"
/clone_1lb="NIH_MGC_85"
/tissue_type="Lymphoma, cell line"
/note="Organ: Lymph. Vector: pCMV-SPORT6; Site_1: NCI;
Site_2: Salt. Cloned unidirectionally; Oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

```

BASE COUNT

189 a 277 c 289 g 143 t

ORIGIN

alignment_scores: 82.50 Length: 134
 Quality: 1.375 Gaps: 6
 Percent Similarity: 44.776 Percent Identity: 28.358

alignment_block:

US-09-471-276-831 x BG177264/rev ..

Align seg 1/1 to reverse of: BG177264 from: 1 to: 898

```

5 ValValaPheLeuLeuLeuTrpGIyValThrcGIyProValThrGInAl 21
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
805 GTTGTGTTTCTGCTGTTTGGGCTGGCGGCTGCTGTG..... 761
21 aAlaIlePheTyrgIuThrcIn**SerLeuTrpAlaGInSerGIuHis* 38
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
760 .....GCTTGTGCGCG..... 749
38 **LeuLysThrLeuGlyGInCysAspAlaAspValProGIyProGIy 54
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
748 .....GGTGGCTGTTGGCTTGTCCGGT 725
55 AspSerArg.....LeuProAlaValGInGInIuTrpGIyAlaGIn. 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
724 CGGTCCACCATGTTAATTTGGCCCAATTCGCCGAGGAGGTGACCGCT 675
68 .....GIuProValHisLeuAspSerP 75
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
674 TCCGTGCTGGGTCTTGACCCACCCCAACCTGCCAAGTTCCAAAC 625
75 roAlaIleLysHisGInPheLeuLeuThr.GIuAspThrGInGlyArgTy 91
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
624 AACCTGCCCCGCCAGATATGAGCTGACTTGGCCACACTAGCGGAATACC 575
91 rArgCysArgSerGIuLeuSerThrGIyTrp**GIuLeuSerLysLeuL 108
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
574 CCAAGGCCCTGAGAGAGTGCCTGGG.....CTGCCAGCACTGC 534
108 euGIuLeuThrGIyProLysValLeuAlaLacysSerLeuAlaLeuAspGIy 124
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
533 AGGAGGACACAGGGGCCACACCCCGCAGATCCAGCTGCATGCTGCC 484
seq_name: gb_est1:AL553624

```

```

seq_documentation_block: 591 bp mRNA EST 16-FEB-2001
LOCUS AL553624 LTI_NFL006_P12 Homo sapiens cDNA clone CS0D1078Y011 5
DEFINITION prime, mRNA sequence.
ACCESSION AL553624
VERSION AL553624.1 GI:12893629
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia: Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 591)
LI,M.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

```

Location/Qualifiers

```

1..591
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1078Y011"
/clone_1lb="LTI_NFL006_P12"
/tissue_type="placenta"

```

BASE COUNT

189 a 277 c 289 g 143 t

/note:"Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a Noli-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@life.com URL: http://fulllength.invitrogen.com"

BASE COUNT 124 a 192 c 161 g 108 t 6 others
ORIGIN

alignment_scores:
Quality: 82.00 Length: 122
Ratio: 1.491 Gaps: 6
Percent Similarity: 45.082 Percent Identity: 31.148

alignment_block:
US-09-471-276-831 x AL553624 ..

Align seg 1/1 to: AL553624 from: 1 to: 591

```

2   SerMetLeuValValPhe.....LeuLeuLeuTrpGlyValThrTrpG1 16
   ::::::::::: ::::: ::::: ::::: ::::: :::::
27  GCCATGATCCCGCCACCTTCACGCGCTGCTGCTGCGCGCTGAGCTCGG 76
   ||||| ::::: ::::: ::::: ::::: :::::
16  yProvalThrGluAlaAlaIlePheTyrgIuThrGln**SerLeuTrpA 33
   ||||| ::::: ::::: ::::: ::::: :::::
77  CCCCAGAGACCCACATGACGAGCGCCCTCCCAACCCACCTCTGGG 126
   ||||| ::::: ::::: ::::: ::::: :::::
33  IAGIuserGluHis**LeuLystrLeuGlyGln...CysAspAlaasp 48
   ||||| ::::: ::::: ::::: ::::: :::::
127 CTGAGCCACGCTGTGATCAG...CTGGGGAACTGTGTGACATCTG 172
   ||||| ::::: ::::: ::::: ::::: :::::
49  ValProGlyProProGlyAspSerArgLeuProAla..... 60
   ||||| ::::: ::::: ::::: ::::: :::::
173 GTCTCAGCGGACCTCGAGCGCTCGGAGTACGCTGATGAACAGGAA 222
   ||||| ::::: ::::: ::::: ::::: :::::
61  .....ValGInGluTrpGlyA 66
   ||||| ::::: ::::: ::::: ::::: :::::
223 GCCCAGACCCCTKGCACATACATKMSCCACTTTATCCCAAGAACAGAGC. 271
   ||||| ::::: ::::: ::::: ::::: :::::
66  IAGInGluProValHisLeuAspSerProAlaIleLyshIsGlnPheLeu 82
   ||||| ::::: ::::: ::::: ::::: :::::
272  .....CAGATTCTCCATCGCGCGATC 292
   ||||| ::::: ::::: ::::: ::::: :::::
83  LeuThrGlyAspThrGlnGlyArgTyrgCysArgSerGlyLeuSerT 99
   ::::: ::::: ::::: ::::: :::::
293 CATGACAGACGATATGACGAGAGATACCGCTGTACTATGCGACCCCTG 342
   ||||| ::::: ::::: ::::: ::::: :::::
99  hrGlyTrp**Gln 103
   ::::: ::::: ::::: ::::: :::::
343 TAGCGTGTGACAG 356
   ||||| ::::: ::::: ::::: ::::: :::::

```

seq_name: gb_est1:AL110419

seq_documentation_block:
LOCUS AL110419 760 bp mRNA EST 29-FEB-2000
DEFINITION DKFZP434L1031.F1 434 (synonym: hles3) Homo sapiens cDNA clone
DKEZP434L1031 5', mRNA sequence.

ACCESSION AL110419
VERSION AL110419.1 GI:5866027
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 760)
AUTHORS Blum,H., Bauer Sachs,S., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE EST (Blum, et al.)
JOURNAL Unpublished (1999)

COMMENT

Contact: Blum H
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone From S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by LMU (Ludwig Maximilians University,
Munich/Germany) within the cDNA sequencing consortium of the German
Genome Project.

FEATURES

source

No sl sequence available
This clone (DKFZP434L1031) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1..760

/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/clone:"DKFZP434L1031"
/clone_lib:"434 (synonym: hles3)"
/issue_type:"Testis"
/dev_stage:"adult"
/lab_host:"DH10B"
/note:"Vector: pSPORT1; Site_1: NotI; Site_2: SalI"

BASE COUNT 185 a 220 c 184 g 171 t
ORIGIN

alignment_scores:
Quality: 82.00 Length: 117
Ratio: 1.242 Gaps: 3
Percent Similarity: 56.410 Percent Identity: 32.479

alignment_block:
US-09-471-276-831 x AL110419 ..

Align seg 1/1 to: AL110419 from: 1 to: 760

```

2   SerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyProva 18
   ::::::::::: ::::: ::::: ::::: ::::: :::::
139 ACCATGCTGAAGACATTCACGTGTTTCTGCTGCTTGCATTCGAGACTGT 188
   ::::: ::::: ::::: ::::: :::::
18  IThrGluAlaAlaIlePheTyrgIuThrGln**SerLeuTrpAlaGln 35
   ::::: ::::: ::::: ::::: :::::
189 GGGTATGACATGATAGTATGATGACCTCAACCGAGTTGTGATAGACT 238
   ||||| ::::: ::::: ::::: ::::: :::::
35  ergLHis**LeuLystrLeuGlyGlnCysAspAlaaspValProGly 51
   ||||| ::::: ::::: ::::: ::::: :::::
239 CCAACTACCCCA. GCCCCTTGGGAGAACATCACTCAGCTTGTGTCGCAAG 287
   ||||| ::::: ::::: ::::: ::::: :::::
52  ProProGlyAsp...SerArgLeuProAlaValGInGluTrpGlyAlaG1 67
   ||||| ::::: ::::: ::::: ::::: :::::
288 CCCCTTCGATATCAAGCAAGCTTCGCTGCTGACAGGATGAACAGACA 337
   ||||| ::::: ::::: ::::: ::::: :::::
67  nGlu.....ProValHisLeuAspSerProAlaIleLyshIsGln. 80
   ::::: ::::: ::::: ::::: :::::
338 TGACCTGATCCGCCCTTCCCAACAAGACTTCCCAAGTTTCTATTCCTATA 387
   ||||| ::::: ::::: ::::: ::::: :::::
81  PheLeuLeuThrGlyAspThrGlnGlyArgTyrgCysArgSerGlyLe 97
   ||||| ::::: ::::: ::::: ::::: :::::
388 GGTCCCTTACTGAGTCCATGCGAGGCTTTTACCGGTGCTGCTACTGAA 437
   ||||| ::::: ::::: ::::: ::::: :::::
97  uSerThrGlyTrp**GlnLeuSerLyshIsGlnGluGlnLeuThrGlyPro 113
   ::::: ::::: ::::: ::::: :::::
438 GGAGACAGCGCTGCTCAAAAGCCCACTAAAGTTCTAGAGTTGAGCGCACCA 486
   ||||| ::::: ::::: ::::: ::::: :::::

```

seq_name: gb_est2:BI331839

seq_documentation_block:
LOCUS BI331839 919 bp mRNA EST 30-JUL-2001
DEFINITION 602982517F1 NCL_CGAP_L19 Mus musculus cDNA clone IMAGE:5135141 5',
mRNA sequence.

ACCESSION BI331839
VERSION BI331839.1 GI:15016496

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 919)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LHAM11332 row: b column: 06
High quality sequence stop: 686.

FEATURES
SOURCE location/Qualifiers
1..919
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5135141"
/clone_lib="NCI CGAP L19"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6, Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."
BASE COUNT 215 a 245 c 279 g 180 t
ORIGIN

alignment_scores:
Quality: 82.00 Length: 86
Ratio: 1.907 Gaps: 4
Percent Similarity: 50.000 Percent Identity: 30.233

alignment_block:
US-09-471-276-831 x B1331839/rev ..

Align seg 1/1 to reverse of: B1331839 from: 1 to: 919

```

43 GlyGlnCysAspAlaAsp.....ValProG1 51
|||||
917 GGAATTGGCCCAAGCAGCTGCAAGAGTACCACCTCCAGTCCCGAG 868
51 YProProGlyAspSerArgLeuProAlaValGlnGluTrpGlyAlaGlnG 68
|||||
867 GGCACACAGCTTCTGCGCCGCTGCGCTGGC...CGACATGGCATGCGCAGC 821
68 IuProValHisLeuAspSerProAlaIleLeuHisGlnPheLeuLeuThr 84
|||||
820 CGCCTTACCATCCACAGCCCTGGA.....ACCTGGCG 786
85 GlyAspThrGlnGlnGlyArgGlyArgSerGlyLeuSerThrGly.. 100
|||||
785 GGACCCACAGCTGCGCGCTTACAGTCCCGGGGTGGCTCAATTCTGGGG 736
101 ....TTP**GlnLeuSerLysLeuLeuGluLeuThrGlyProLysVal 116
|||||
735 AACCTGGCACAGGACCGTGTATCTTGCACCTCAGTGTCCACCGTGAA 686
116 euAlaCys 118
|||||
685 CAGCATGC 678
seq_name: gb_est2:BF347835
seq_documentation_block:

```

LOCUS BF347835 1101 bp mRNA EST 22-NOV-2000
DEFINITION 602022972P1 NCI CGAP_Brn67 Homo sapiens CDNA clone IMAGE:4158569
ACCESSION BF347835
VERSION BF347835.1 GI:11295430
KEYWORDS 5', mRNA sequence.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1101)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LHAM9434 row: o column: 18
High quality sequence stop: 573.

FEATURES
SOURCE location/Qualifiers
1..1101
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4158569"
/clone_lib="NCI CGAP_Brn67"
/tissue_type="anaplastic oligodendroglioma with 1p/19q loss"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: brain; Vector: pCMV-SPORT6, Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."
BASE COUNT 369 a 292 c 216 g 224 t
ORIGIN

alignment_scores:
Quality: 82.00 Length: 136
Ratio: 1.206 Gaps: 7
Percent Similarity: 50.000 Percent Identity: 27.206

alignment_block:
US-09-471-276-831 x BF347835/rev ..

Align seg 1/1 to reverse of: BF347835 from: 1 to: 1101

```

13 ValThrTrpGlyProValIhrGlnAlaIlePheTyrgLutHrcIn** 29
|||||
569 TTGACCTGGGCGGAGTCCAGTCAAGTCACTGTGATATATGCTGCT 520
29 *SerLeu.....TrpAlaGluSerGluHis**LeuLysThrL 42
|||||
519 GCGCTTATTAATACCTAGTTCCTTTACAGTACGAGAGAGTGAAGTGT 470
42 euGlyGln.....CysAspAlaAspValProGlyPro... 52
|||||
469 TGGACGCGCTCGGCACAGTCTGCTGCCAGTGAAGTCACTGCTCAGGT 420
53 .....ProGlyAspSerArgLeuProAlaValGlnGlu..TrpLya 66
|||||
419 GCACTGCTCCAGGAGGCTCTGAGAGGTCCTCCCTTCGACAGAGCTGGAG 370
66 IacGlnIuProValHisLeuAspSerProAlaIleLysHisGlnPheLeu 82
|||||
369 CTTCGCCACAGCGGCGGACGTTCACTCAGTCAAGTCAATTCGTGCTGGG 320
83 LeuThrGlnAspThrGlnGlyArgTyArg..... 92

```

```

319 TCAGCAGAGTAGTACTCAATAGAGTCCGAGCAGCGCGCTTGCCATGCCCT 270
93 .....CysArgSerGlyLeuSerThrGlyT 101
269 GACACCTTGCGCGCTGACGACGACCTGTAGTCCCTGCT.....ACTGCTT 226
101 rp**GlnLeu.....SerLysLeuLeuGlnLeuThrGlyProLysVal 115
225 GCGCTAGCAGCTGTGTGGGAGACCTGCTTGTCTTCTCCCGGAGACTG 176
116 LeuAla 117
175 ATCGCA 170

```

seq_name: gb_est1:AL597254

seq_documentation_block:

LOCUS AL597254 419 bp mRNA EST 14-AUG-2001
DEFINITION DKF2P313B1512.F1 313 (synonym: hicc2) Homo sapiens cDNA clone
DKF2P313B1512 5', mRNA sequence.

ACCESSION AL597254
VERSION AL597254.1 GI:15156094

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 419)
Ansoerge, W., Winkner, U., Mewes, W., Well, B. and Wiemann, S.

AUTHORS EST (Ansoerge, W., Winkner, U., Mewes, W., Well, B. and Wiemann, S.)
TITLE Unpublished (1999)
JOURNAL Contact: Ansoerge W

COMMENT

MIPS

Am Klopferstr. 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

Sequenced by EMBL (European Molecular Biology Laboratories,

Heidelberg/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No sl sequence available.

This clone (DKF2P313B1512) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
source

1..419
Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="DKF2P313B1512"

/clone_lib="313 (synonym: hicc2)"

/dev_stage="adult"

/lab_host="DH10B"

/note="Vector: pT7Blue; Site_1: SfiI; Site_2: SfiI; CDNA-collection"

BASE COUNT 86 a 145 c 122 g 66 t

ORIGIN

alignment_scores:

Quality: 81.50 Length: 133

Ratio: 1.482 Gaps: 6

Percent Similarity: 41.353 Percent Identity: 27.820

alignment_block:

US-09-471-276-831 x AL597254 ..

Align seg 1/1 to: AL597254 from: 1 to: 419

1 MetSerMetLeuValValPheLeuLeuLeuThrGlyValThrTrpGlyPr 17

||||| :||||| :||||| :||||| :||||| :||||| :|||||

82 ATGACCGCCCGCTGACAGCCCTGCTGCGCTGCGCTGAGTGGGCC 131

seq_name: gb_est2:BG257260

seq_documentation_block:

LOCUS BG257260 862 bp mRNA EST 13-FEB-2001
DEFINITION 60237975F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4508562 5',
mRNA sequence.

ACCESSION BG257260
VERSION BG257260.1 GI:12767076

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 862)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: ILAMI0387 row: f column: 19

High quality sequence stop: 661.

Location/Qualifiers

1..862

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4508562"

/clone_lib="NIH_MGC_92"

/tissue_type="embryonal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: testis; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 2.5 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC library."

BASE COUNT

192 a 280 c 234 g 156 t

ORIGIN

alignment_scores:
 Quality: 81.50 Length: 82
 Ratio: 2.038 Gaps: 5
 Percent Similarity: 48.780 Percent Identity: 36.585

alignment_block:
 US-09-471-276-831 x BG257260 ..

Align seg 1/1 to: BG257260 from: 1 to: 862

```

46 Aspa1aAspValPro.....GlyProProGI 54
||||| :|||
279 GAGCGCATGATCCCGCCTTCAGCGCTGCTGCTGCGGAGCCCTCG 328
54 yAspSerAArgLeuProAla.ValGlnGlu.....TPPGly 65
||||| :|||
329 AGGCTCGGAGGATGACGCTGATTAAGAGGAAAGCCAGCACCCTGGAC 378
66 AlaGlnGluProValHisLeuAspSerProAlaIleLysHisGlnPhe.. 81
||||| :|||
379 AGACAGAACCCACTG.....GAGCCCAAGACAGAGCCGAGATTCTC 419
82 .....LeuLeuThrGlyAspThrGlnGlyArgTyrArgCysArgSerC 96
||||| :|||
420 CATCCCATGCTACAGAGGACTATGACAGAGATACCGCTGTACTATC 469
96 LyuSerThrGlyTyr***GlnLeuSerLysLeuGluLeu 110
||||| :|||
470 GAGCCCTGTAGCTGTGTCACAGCCAGTACGCCCTGTGAGCTG 513

```

seq_name: gb_est2:BF195074

seq_documentation_block:
 LOCUS BF195074 581 bp mRNA EST 03-NOV-2000
 DEFINITION 7094e12.x1 NCI-CGAP.Ov18 Homo sapiens cDNA clone IMAGE:3643895 3'
 similar to contains MER22.b3 MER22 repetitive element ;, mRNA
 sequence.

ACCESSION BF195074 GI:11081567
 VERSION BF195074.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 581)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
 Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 High quality sequence stop: 526.

FEATURES

Location/Qualifiers
 1..581
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3643895"
 /clone_lib="NCI-CGAP.Ov18"
 /tissue_type="fibrobloma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pT733-Pac (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTCACATCTGAACTGGAGCGCGCGGACATTTTATTTTATTTT 3'];
 double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 92 a 174 c 213 g 102 t
 ORIGIN

alignment_scores:
 Quality: 81.00 Length: 117
 Ratio: 1.884 Gaps: 4
 Percent Similarity: 36.752 Percent Identity: 25.641

alignment_block:
 US-09-471-276-831 x BF195074 ..

Align seg 1/1 to: BF195074 from: 1 to: 581

```

15 TRPGlyProValThrGlnAlaIlePheThrGln***SerLe 31
||||| :|||
30 TGGGTCCTCACTGTACAGAGGCGCAGTTCACAGAAAGGGTTCACTG 79
31 u...TPAlaGluSerGluHis**LeuYsThrLeuGlyGlnCysAspa 47
||||| :|||
80 GCGCTGGCGAGCCAGCCACAGAGAGAGAGAGAGAGAGAGAGAGAG 105
47 laAspValPro.....G 50
||||| :|||
106 CTGATGTTCACACTGTGGCAGGTGGAGAGATGCCGAGCTCGTGAGCCCA 155
50 .....G 50
156 GTGCCAGAGGAGCTGGAGACGCCGAGGGGCGCGGAGCCCTCCT 205
51 .....GlyProProGlyAspSerArgLeuP 59
||||| :|||
206 AGTCACCCAGAGGAGGCGCTGTGTCACCCCAAGAGCTGTGGAGCTC 255
59 roAlaValGlnGluTyrGlyAlaGlnGluProValHisLeuAspSerPro 75
||||| :|||
256 CTGCGGTGCGCCCGCAGAGTCACTCCGAGCCCGCCAGCAGCAGAGCCCT 305
76 AlaIle.....LysHisGlnPheLeuLeuThrGlyAspTh 87
||||| :|||
306 GCGTCTGAGTACTGCGGCGCATTCACTACAGTGTACTAGAGGTGCCAC 355
87 r 87
356 G 356

```

1
2
3


```

85      39      44.8      301      21      AAY57945      Human transmembran
86      39      44.8      301      22      AAM39049      Human polypeptide
87      39      44.8      301      22      AAU12254      Human PRO4343 poly
88      39      44.8      314      22      AAM40835      Human polypeptide
89      39      44.8      332      21      AAY82703      Tick derived cyste
90      39      44.8      344      18      AAM26767      Human chemokine re
91      39      44.8      344      19      AAM23957      Amino acid sequenc
92      39      44.8      356      19      AAM48087      Human macrophage/d
93      39      44.8      363      20      AAM94654      G-protein coupled
94      39      44.8      363      22      AAU04379      Human G-protein co
95      39      44.8      387      21      AAY90637      Human G-protein-co
96      39      44.8      387      21      AAY90672      Human mutant G-pro
97      39      44.8      457      22      AAU03813      G-protein-coupled
98      39      44.8      473      22      AAG81483      S. epidermidis ope
99      39      44.8      497      22      AAG90137      C glutamicum prote
100     39      44.8      499      22      AAG90588      C glutamicum prote

```

ALIGNMENTS

```

RESULT 1
ID      AAY64670 standard; Protein: 126 AA.
XX
AC      AAY64670;
XX
DT      01-FEB-2000 (first entry)
XX
DE      Human 5' EST related polypeptide SEQ ID NO:831.
XX
KM      Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KM      gene therapy; chromosome mapping; upstream regulatory sequence;
KM      forensic; location; development; protein synthesis; stability;
KM      regulation; identification.
XX
OS      Homo sapiens.
XX
PN      WO953051-A2.
XX
PD      21-OCT-1999.
XX
PF      09-APR-1999; 99WO-1B00712.
XX
PR      09-APR-1998; 98US-0057719.
PR      28-APR-1998; 98US-0069047.
XX
PA      (GEST ) GENSET.
XX
PI      Dumas MLine Edwards J, Duclert A, Giordano J;
XX
DR      WPI: 2000-038446/03.
XX
DR      N-PSDB; AA242284.
XX
PT      Novel secreted protein 5' expressed sequence tag sequences used in
PT      diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS      Claim 3; Page 603; 837pp; English.
XX
CC      AA242285 to AA243075 represent novel 5' expressed sequence tag (EST)
CC      sequences, corresponding to human secreted proteins. AAY64651 to
CC      AAY5548 represent the EST-related proteins corresponding to AA242285 to
CC      AA243082. The 5' ESTs can be used for producing secreted human gene
CC      products. They can be used to identify and isolate 5' untranslated
CC      regions (UTRs) and upstream regulatory regions which control the
CC      location, development stage, rate, and quantity of protein synthesis, as
CC      well as stability of mRNA. The ESTs are also useful as probes for
CC      chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC      also be used in forensic procedures to identify individuals, or in
CC      diagnostic procedures to identify individuals having genetic diseases
CC      resulting from abnormal gene expression. The products may also be used in
CC      gene therapy protocols. The nucleic acids encoding signal peptides can be
CC      used for directing extracellular secretion of a polypeptide or the

```

```

CC      Insertion of a polypeptide into a membrane, or importing a polypeptide
CC      into a cell. The proteins encoded by the EST sequences may be useful in
CC      treating a variety of human conditions. Secreted proteins have
CC      therapeutic value, and the identification of new secreted proteins is
CC      valuable. AA242249 to AA242264 and AAY64644 to AAY64650 represent
CC      sequences used in the exemplification of the present invention.
XX
SQ      Sequence 126 AA:

```

```

Query Match      100.0%; Score 87; DB 21; Length 126;
Best Local Similarity 100.0%; Pred. No. 2,3e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      1 MSMLVFFLLMGVTCG 16
      |||||
DB      1 MSMLVFFLLMGVTCG 16

```

```

RESULT 2
ID      AAB48724 standard; Protein: 512 AA.
XX
AC      AAB48724;
XX
DT      09-MAR-2001 (first entry)
XX
DE      Mouse liver growth hormone-induced clone 5 ORF #1, SEQ ID NO:8.
XX
KM      Mouse; growth hormone; GH regulatable gene; liver pathology; hypertrophy;
KM      hepatocellular lesion; hyperplasia; altered expression level; clone 5;
KM      diagnostic marker; gigantism; acromegaly; diabetes; hepatotropic;
KM      transgenic animal; drug screening; drug discovery; murine;
KM      cytoplasmic protein; open reading frame; ORF.
XX
OS      Mus sp.
XX
PN      WO200066787-A2.
XX
PD      09-NOV-2000.
XX
PF      05-MAY-2000; 2000WO-US12366.
XX
PR      05-MAY-1999; 99US-0132663.
XX
PA      (UYOH-) UNIV OHIO.
XX
PI      Kopschick JU, Tlong J;
XX
DR      WPI: 2001-007239/01.
XX
DR      N-PSDB; AAC87257.
XX
PT      Diagnosing abnormal levels of growth hormone activity in liver
PT      comprising assaying growth transcriptional activity and protein
PT      expression level of hormone-regulatable liver genes, as diagnostic
PT      markers of liver pathology
XX
PS      Claim 2; Page 53; 65pp; English.
XX
CC      The invention relates to a method of diagnosing abnormal levels of
CC      growth hormone (GH) activity in the liver, or predicting a change in the
CC      condition of the liver in response to abnormal GH activity. The method
CC      involves correlating the level of expression of certain specific genes
CC      with the level of GH activity in the liver, or with an expected change
CC      in the condition of the liver as the result of GH activity. Excessive GH
CC      activity in the liver is thought to be deleterious to health, causing an
CC      increase in liver size as a consequence of both hyperplasia and
CC      hepatocyte hypertrophy, and hepatocellular lesions which progress with
CC      age. Studies in transgenic mice which express high levels of bovine
CC      growth hormone identified a number of genes whose expression in the
CC      liver is altered by high GH levels. The genes which are upregulated are
CC      those encoding alpha-fetoprotein, corticosteroid binding globulin,
CC      fetuin, rab8-interacting protein, paraoxonase-3, cytochrome P45011A,

```


CC artery disease. The present sequence is the human M019 protein.
 XX
 SQ Sequence 85 AA:

Query Match 55.2%; Score 48; DB 22; Length 85;
 Best Local Similarity 72.7%; Pred. No. 5;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 6 VFLLMGVTWG 16
 |||||
 Db 12 VFLLMGVTWG 22

RESULT 5
 AAB48740
 ID AAB48740 standard: Protein: 342 AA.

XX AAB48740;
 XX

DT 09-MAR-2001 (first entry)

XX Mouse liver growth hormone-induced clone 45 ORF #1, SEQ ID NO:10.

DE Mouse; growth hormone; GH regulatable gene; liver pathology: hypertrophy;
 KW hepatocellular lesion; hyperplasia; altered expression level; clone 45;
 KW diagnostic marker; gigantism; acromegaly; diabetes; hepatocytic;
 KW transgenic animal; drug screening; drug discovery; murine;
 KW membrane bound protein; secreted protein; HLH; open reading frame; ORF;
 KM Myc-type helix-loop-helix dimerisation domain.
 OS Mus sp.

XX WO200066787-A2.
 XX

PD 09-NOV-2000.

XX 05-MAY-2000; 2000WO-US12366.
 XX

XX 05-MAY-1999; 99US-0132663.
 XX

XX (UYOH-) UNIV OHIO.
 XX

PI Kopeck JJ, Tjong J;
 XX

DR WPI: 2001-007239/01.
 DR N-PSDB: AAC87258.

XX Diagnosing abnormal levels of growth hormone activity in liver
 PT comprising assaying growth transcriptional activity and protein
 PT expression level of hormone-regulatable liver genes, as diagnostic
 PT markers of liver pathology -
 XX

PS Example 2; Page 57; 65pp: English.

XX The invention relates to a method of diagnosing abnormal levels of
 CC growth hormone (GH) activity in the liver, or predicting a change in the
 CC condition of the liver in response to abnormal GH activity. The method
 CC involves correlating the level of expression of certain specific genes
 CC with the level of GH activity in the liver, or with an expected change
 CC in the condition of the liver as the result of GH activity. Excessive GH
 CC activity in liver size is thought to be deleterious to health, causing an
 CC increase in liver size as a consequence of both hyperplasia and
 CC hepatocyte hypertrophy, and hepatocellular lesions which progress with
 CC age. Studies in transgenic mice which express high levels of bovine
 CC growth hormone identified a number of genes whose expression in the
 CC liver is altered by high GH levels. The genes which are upregulated are
 CC those encoding alpha-fetoprotein, corticosteroid binding globulin,
 CC fetuin, rat8-interacting protein, paraoxonase-3, cytochrome P450I1A,
 CC S-2 hydroxycid oxidase, interferon alpha/beta receptor, growth hormone
 CC receptor, proteasome z-subunit, and coagulation factor V. Two novel
 CC genes, clone 5 (AAC87257) and clone 45 (AAC87258) are also upregulated
 CC in response to abnormally high GH levels. Conversely, expression of the

CC gene encoding 3-beta-hydroxysteroid dehydrogenase/delta-5-delta-4
 CC isomerase appears to be downregulated. The invention also relates to
 CC transgenic nonhuman mammals comprising a GH-induced transgene which
 CC exhibit or have a propensity to develop a liver pathology; an assay for
 CC drugs which inhibit the development of, or which treat a liver pathology,
 CC comprising administering the drug to the transgenic animal; and
 CC preventing or treating a liver pathology in a patient comprising
 CC administering a drug which inhibits the expression of a GH-induced gene.
 CC The method of the invention is used for diagnosing abnormal levels of GH
 CC activity in the liver or predicting a change in the condition of the
 CC liver in response to abnormal levels of GH activity. The GH-regulatable
 CC liver genes and proteins are useful as diagnostic markers of liver
 CC pathology. Assays for the expression of these genes is useful for the
 CC diagnosis of liver pathologies associated with gigantism or acromegaly or
 CC with diabetes, as other causative agents may act directly or indirectly
 CC upon the same genes. The present sequence represents the protein encoded
 CC by the longest open reading frame of the novel mouse liver cDNA clone 45.
 CC The protein is thought to be membrane bound and/or secreted, and has a
 CC Myc-type helix-loop-helix dimerisation domain.
 XX

XX Sequence 342 AA:

Query Match 54.0%; Score 47; DB 22; Length 342;
 Best Local Similarity 64.3%; Pred. No. 30;
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 MSNLVYFLLMGVT 14
 ||:| |||||
 Db 1 MSNLVYFLLMGVT 14

RESULT 6
 AAG92182
 ID AAG92182 standard: Protein: 148 AA.

XX AAG92182;
 XX

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 5936.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 KW

XX Corynebacterium glutamicum.
 OS

PN EP1108790-A2.
 PN

PD 20-JUN-2001.
 PD

PF 18-DEC-2000; 2000EP-0127688.
 PF

XX 16-DEC-1999; 99JP-0377484.
 XX

PR 07-APR-2000; 2000JP-0159162.
 PR

PR 03-AUG-2000; 2000JP-0280988.
 XX

PA (KYOMA) KYOMA HAKKO KOGYO KK.
 PA

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX

DR WPI: 2001-376931/40.
 DR N-PSDB: AAH67401.

XX Novel polynucleotides derived from Coryneform bacterie, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT -

PS Claim 17; SEQ ID NO: 5936; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
 CC

CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
CC from Corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

CC Sequence 148 AA;

Query Match 50.6%; Score 44; DB 22; Length 148;
Best Local Similarity 66.7%; Pred. No. 33;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVPELLMG 12
| : : | | | | | | | |
Db 98 mgpvmvfillwg 109

RESULT 7

AAM41807
ID AAM41807 standard; Protein: 168 AA.

AC AAM41807;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6738.

XX Human; noctropic; immunosuppressant; cytosolic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

DR MPI: 2001-442253/47.
DR N-PSDB: AAI60963.

PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 6738; 10078pp; English.
XX

CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with noctropic.
CC immunosuppressant and cytosolic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

CC Sequence 168 AA;

Query Match 50.6%; Score 44; DB 22; Length 168;
Best Local Similarity 53.8%; Pred. No. 37;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 MLVFPLLMGVTW 15
| : | | | | | | | |
Db 37 laivllllwgapw 49

RESULT 8

AAG28190
ID AAG28190 standard; Protein: 200 AA.

AC AAG28190;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 33316.

XX Arabidopsis thaliana protein fragment
KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148368.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 50.6%; Score 44; DB 21; Length 200;
Best Local Similarity 60.0%; Pred. No. 45;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 SMLVFLLMGVTWG 16
| | | | | | | |
Db 171 sflasflgwtflg 185

RESULT 9
AAC28189
ID AAC28189 standard; Protein: 228 AA.
XX
AC AAC28189;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment seq ID NO: 33315.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KM hydrolisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137702.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142053.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145194.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.

PR	02-AUG-1999;	9905-0146386;
PR	02-AUG-1999;	9905-0146388;
PR	02-AUG-1999;	9905-0146389;
PR	03-AUG-1999;	9905-0147038;
PR	04-AUG-1999;	9905-0147204;
PR	04-AUG-1999;	9905-0147302;
PR	05-AUG-1999;	9905-0147192;
PR	05-AUG-1999;	9905-0147260;
PR	06-AUG-1999;	9905-0147436;
PR	06-AUG-1999;	9905-0147410;
PR	09-AUG-1999;	9905-0147493;
PR	09-AUG-1999;	9905-0147435;
PR	10-AUG-1999;	9905-0148171;
PR	11-AUG-1999;	9905-0148341;
PR	12-AUG-1999;	9905-0148341;
PR	13-AUG-1999;	9905-0148565;
PR	13-AUG-1999;	9905-0148684;
PR	16-AUG-1999;	9905-0149368;
PR	17-AUG-1999;	9905-0149175;
PR	18-AUG-1999;	9905-0149426;
PR	20-AUG-1999;	9905-0149722;
PR	20-AUG-1999;	9905-0149723;
PR	20-AUG-1999;	9905-0149722;
PR	20-AUG-1999;	9905-0149929;
PR	23-AUG-1999;	9905-0149902;
PR	23-AUG-1999;	9905-0149930;
PR	25-AUG-1999;	9905-0150566;
PR	26-AUG-1999;	9905-0150584;
PR	26-AUG-1999;	9905-0150864;
PR	27-AUG-1999;	9905-0151065;
PR	27-AUG-1999;	9905-0151066;
PR	27-AUG-1999;	9905-0151080;
PR	30-AUG-1999;	9905-0151303;
PR	31-AUG-1999;	9905-0151303;
PR	01-SEP-1999;	9905-0151530;
PR	01-SEP-1999;	9905-0151530;
PR	07-SEP-1999;	9905-0152363;
PR	10-SEP-1999;	9905-0153700;
PR	13-SEP-1999;	9905-0153758;
PR	15-SEP-1999;	9905-0154018;
PR	16-SEP-1999;	9905-0154079;
PR	20-SEP-1999;	9905-0154579;
PR	22-SEP-1999;	9905-0155133;
PR	23-SEP-1999;	9905-0155133;
PR	24-SEP-1999;	9905-0155659;
PR	28-SEP-1999;	9905-0156458;
PR	29-SEP-1999;	9905-0156596;
PR	05-OCT-1999;	9905-0157117;
PR	05-OCT-1999;	9905-0157153;
PR	06-OCT-1999;	9905-0157573;
PR	07-OCT-1999;	9905-0158025;
PR	08-OCT-1999;	9905-0158629;
PR	12-OCT-1999;	9905-0158332;
PR	13-OCT-1999;	9905-0158369;
PR	13-OCT-1999;	9905-0158293;
PR	13-OCT-1999;	9905-0159294;
PR	13-OCT-1999;	9905-0159295;
PR	14-OCT-1999;	9905-0158329;
PR	14-OCT-1999;	9905-0159330;
PR	14-OCT-1999;	9905-0159331;
PR	14-OCT-1999;	9905-0159637;
PR	14-OCT-1999;	9905-0159638;
PR	18-OCT-1999;	9905-0159584;
PR	21-OCT-1999;	9905-0160741;
PR	21-OCT-1999;	9905-0160767;
PR	21-OCT-1999;	9905-0160768;
PR	21-OCT-1999;	9905-0160770;
PR	21-OCT-1999;	9905-0160815;
PR	22-OCT-1999;	9905-0160815;
PR	22-OCT-1999;	9905-0160980;
PR	22-OCT-1999;	9905-0160981;
PR	25-OCT-1999;	9905-0161404;
PR	25-OCT-1999;	9905-0161405;
PR	25-OCT-1999;	9905-0161406;
PR	26-OCT-1999;	9905-0161350;
PR	26-OCT-1999;	9905-0161369;

PR	26-OCT-1999;	98US-0161361.
PR	28-OCT-1999;	98US-0161920.
PR	28-OCT-1999;	98US-016192.
PR	28-OCT-1999;	98US-016192.
PR	28-OCT-1999;	98US-016193.
PR	29-OCT-1999;	98US-0162142.
Query Match		
Best Local Similarity		50.6%; Score 44; DB 21; Length 228;
Matches		9; Conservative
OY	2 SMLVFLIMGVTWG 16	
	:	
Db	199 sflastlllgwtlig 213	
RESULT 10		
AAV13399		
ID	AAV13399 standard; Protein; 280 AA.	
XX	AAV13399;	
XX		
XX	25-JUN-1999 (first entry)	
DE	Amino acid sequence of protein PRO268.	
XX		
KW	Secreted protein; transmembrane protein; human; enterocolitis;	
KW	Zollinger-Ellison syndrome; gastrointestinal ulceration;	
KW	congenital microvillus atrophy; skin disease; cell growth;	
KW	abnormal keratinocyte differentiation; psoriasis; epithelial cancer;	
KW	Parkinson's disease; Alzheimer's disease; ALS; neuropathy;	
KW	fibrinodermis; dermal scarring; Usher Syndrome; Atrophila areata;	
KW	antithrombotic; wound healing; tissue repair.	
XX		
OS	Homo sapiens.	
XX		
PN	WO9914328-A2.	
XX		
PD	25-MAR-1999.	
XX		
PF	16-SEP-1998;	98MO-US19330.
XX		
PR	25-NOV-1997;	97US-0066840.
PR	17-SEP-1997;	97US-0059113.
PR	17-SEP-1997;	97US-0059115.
PR	17-SEP-1997;	97US-0059117.
PR	17-SEP-1997;	97US-0059119.
PR	17-SEP-1997;	97US-0059121.
PR	17-SEP-1997;	97US-0059122.
PR	17-SEP-1997;	97US-0059184.
PR	18-SEP-1997;	97US-0059263.
PR	18-SEP-1997;	97US-0059266.
PR	15-OCT-1997;	97US-0062125.
PR	17-OCT-1997;	97US-0062285.
PR	17-OCT-1997;	97US-0062287.
PR	21-OCT-1997;	97US-0063486.
PR	24-OCT-1997;	97US-0062814.
PR	24-OCT-1997;	97US-0062816.
PR	24-OCT-1997;	97US-0063045.
PR	24-OCT-1997;	97US-0063120.
PR	24-OCT-1997;	97US-0063121.
PR	24-OCT-1997;	97US-0063128.
PR	27-OCT-1997;	97US-0063329.
PR	27-OCT-1997;	97US-0063327.
PR	28-OCT-1997;	97US-0063541.
PR	28-OCT-1997;	97US-0063542.
PR	28-OCT-1997;	97US-0063544.
PR	28-OCT-1997;	97US-0063550.
PR	28-OCT-1997;	97US-0063564.
PR	29-OCT-1997;	97US-0063435.
PR	29-OCT-1997;	97US-0063704.
PR	29-OCT-1997;	97US-0063733.


```

PR 29-OCT-1997; 97US-0063738.
PR 29-OCT-1997; 97US-0063734.
PR 29-OCT-1997; 97US-0064215.
PR 29-OCT-1997; 97US-0063735.
PR 31-OCT-1997; 97US-0063870.
PR 31-OCT-1997; 97US-0064103.
PR 03-NOV-1997; 97US-0064248.
PR 07-NOV-1997; 97US-0064809.
PR 12-NOV-1997; 97US-0065186.
PR 17-NOV-1997; 97US-0065846.
PR 18-NOV-1997; 97US-0065693.
PR 21-NOV-1997; 97US-0066120.
PR 21-NOV-1997; 97US-0066364.
PR 24-NOV-1997; 97US-0066772.
PR 24-NOV-1997; 97US-0066466.
PR 24-NOV-1997; 97US-0066770.
PR 24-NOV-1997; 97US-0066511.
PR 24-NOV-1997; 97US-0066453.
XX
XX
PA (GENH ) GENENTECH INC.
XX
PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
XX
DR WPI: 1999-229533/19.
DR N-PSDB; AAX52270.
XX
XX PT New isolated human genes and polypeptides used in, e.g. treatment of
PT gastrointestinal ulceration
XX
XX PS Claim 12; Fig 114; 320pp; English.
XX
XX CC AAY13344-403 represent secreted and transmembrane human proteins.
CC The cDNA sequences are obtained from cDNA libraries, prepared from
CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC The encoded polypeptides have specific uses based on their homology to
CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC associated with the preservation and maintenance of gastrointestinal
CC mucosa and the repair of acute and chronic mucosal lesions
CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC ulceration and congenital microvillus atrophy), skin diseases associated
CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC potent effects on cell growth and development, diseases related to
CC growth or survival of nerve cells including Parkinson's disease,
CC Alzheimer's disease, AIDS, neuropathies or cancer. PRO265 can be used as
CC for fibronectin, e.g. for reducing dermal scarring. PRO264 can be used
CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
CC of Usher Syndrome or Atrophila areata; PRO269 can be used as an
CC anti-thrombotic agent; PRO287 polypeptides and portions may have
CC therapeutic applications in wound healing and tissue repair; PRO317 can
CC be used for treating problems of the kidney, uterus, endometrium, blood
CC vessels, or related tissue, e.g. in the heart of genital tract.
XX
XX SQ Sequence 280 AA:

```

```

Query Match 50.6%; Score 44; DB 20; Length 280;
Best Local Similarity 53.8%; Pred. No. 64;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 3 MLVFFLLMGVTM 15
DB 11 IAVIVLLWGPW 23

```

```

RESULT 11
AAB19404
ID AAB19404 standard; Protein: 280 AA.
XX
XX AC AAB19404;
XX
XX DT 06-MAR-2001 (first entry)
XX

```

```

DE Amino acid sequence of a human secreted protein.
XX
XX KW Secreted protein; platelet disorder; stem cell disorder; osteoporosis;
KW osteoarthritis; burn; incision; ulcer; periodontal disease; neuropathy;
KW nervous system disease; bone growth; cosmetic plastic surgery;
KW gut protection; gut regeneration; fibrosis; cancer;
KW bone marrow transplantation.
XX
XX OS Homo sapiens.
XX
XX PN W0200061755-A2.
XX
XX PD 19-OCT-2000.
XX
XX PE 10-APR-2000; 2000WO-US09555.
XX
XX PR 09-APR-1999; 99US-0128574.
XX PR 20-AUG-1999; 99US-0150054.
XX
XX PA (CHIR ) CHIRON CORP.
XX
XX PI Garcia PD:
XX
XX DR WPI: 2000-665133/64.
XX DR N-PSDB; AAC61890.
XX
XX PT Novel secreted human proteins useful for stimulating blood cell
PT generation in patients receiving cancer chemotherapy, treating bone
PT marrow transplantation patients and for healing fractured bones
XX
XX PS Claim 14; Page 69; 74pp; English.
XX
XX CC AAB19393-B19407 represent secreted human proteins. The secreted proteins
CC are useful in assays to determine their biological activities. The
CC proteins can also be used as biomarkers to identify tissues or cell
CC types which express the proteins. The polynucleotide molecules can be
CC used as biomarkers for tissues or chromosomes and to elicit immune
CC responses. The proteins and antibodies are useful in diagnosis and
CC treatment of diseases associated with altered expression of these
CC proteins. The proteins are also useful for prevention or treatment of
CC platelet disorders, stem cell disorders, osteoporosis or osteoarthritis,
CC burns, incisions, ulcers, periodontal diseases, central and peripheral
CC nervous system diseases and neuropathies, for healing fractured bones
CC and to induce cartilage and/or bone growth in cosmetic plastic surgery.
CC The proteins are also useful for gut protection or regeneration, for the
CC treatment of lung or liver fibrosis, for stimulating blood cell
CC generation in patients receiving cancer chemotherapy and for treatment
CC of bone marrow transplantation patients.
XX
XX SQ Sequence 280 AA:

```

```

Query Match 50.6%; Score 44; DB 21; Length 280;
Best Local Similarity 53.8%; Pred. No. 64;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 3 MLVFFLLMGVTM 15
DB 11 IAVIVLLWGPW 23

```

```

RESULT 12
AAB18615
ID AAB18615 standard; Protein: 280 AA.
XX
XX AC AAB18615;
XX
XX DT 15-JAN-2001 (first entry)
XX
XX DE DNA encoding a human thioredoxin family active site molecule MP-4.
XX MP-4; thioredoxin family active site molecule; cellular redox reaction;
XX cardiovascular disease; hypertension; atherosclerosis; valvular disease;
KW

```

KM coronary artery spasm; coronary artery disease; arrhythmia;
 KM myocardial infarction; cardiac hypertrophy; cardiomyopathy;
 KM congestive heart failure; connective tissue disorder; hepatic disorder;
 KM Ehlers-Danlos Syndrome; alcoholic liver disorder; liver cirrhosis;
 KM liver cancer.
 XX Homo sapiens.
 OS
 XX
 FH Key
 FT Peptide
 FT 1.23 Location/Qualifiers
 FT /note= "signal peptide"
 FT 20..25 /note= "N-myristoylation site"
 FT 24..280 Protein
 FT /note= "mature protein"
 FT 25..28 /note= "amidation site"
 FT 28..126 /note= "active-site"
 FT /note= "thioredoxin family active site"
 FT 68..71 /note= "casein kinase II phosphorylation site"
 FT 89..105 /note= "transmembrane domain"
 FT 92..94 /note= "protein kinase C phosphorylation site"
 FT 106..114 /note= "tyrosine kinase phosphorylation site"
 FT 118..120 /note= "protein kinase C phosphorylation site"
 FT 119..121 /note= "casein kinase II phosphorylation site"
 FT 127..130 /note= "casein kinase II phosphorylation site"
 FT 127..129 /note= "protein kinase C phosphorylation site"
 FT 137..161 /note= "transmembrane domain"
 FT 182..206 /note= "transmembrane domain"
 FT 191..196 /note= "N-myristoylation site"
 FT 208..210 /note= "protein kinase C phosphorylation site"
 FT 220..222 /note= "protein kinase C phosphorylation site"
 FT 246..249 /note= "casein kinase II phosphorylation site"
 FT 256..259 /note= "casein kinase II phosphorylation site"
 FT 256..258 /note= "protein kinase C phosphorylation site"
 FT /note= "protein kinase C phosphorylation site"
 FT
 XX WO200053769-A1.
 PN
 XX
 PD 14-SEP-2000.
 XX
 PE 06-MAR-2000; 2000WO-US05725.
 XX
 PR 08-MAR-1999; 99US-0264419.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI khodadoust MM;
 XX
 DR WPI: 2000-594324/56.
 DR N-PSDB; AAA75487, AAA75488.
 XX
 PT Novel nucleic acids encoding MP-4 polypeptides of thioredoxin family
 PT active site molecule useful for treating cardiovascular disorders and
 PT congestive heart failure -
 XX
 PS Claim 8; Fig 1; 109pp; English.
 XX

CC The present sequence represents a polypeptide designated MP-4, which is
 CC a member of thioredoxin family active site molecules. MP-4 is a
 CC regulator of target cellular redox reactions. MP-4 is a
 CC and polypeptides are useful for treating cardiovascular diseases such
 CC as hypertension, atherosclerosis, coronary artery spasm, coronary
 CC artery disease, valvular disease, arrhythmia, myocardial infarction,
 CC cardiac hypertrophy and cardiomyopathies, and congestive heart failure.
 CC polypeptide is also useful for treating connective tissue disorder such
 CC as Ehlers-Danlos Syndrome, or a hepatic disorder such as alcoholic liver
 CC disorder, liver cirrhosis and liver cancer.
 CC
 XX
 SO Sequence 280 AA:
 OY 3 MLVFFLLNGVTM 15
 Db 11 IAVIVIIIVGAPW 23
 Query Match 50.6%; Score 44; DB 21; Length 280;
 Best Local Similarity 53.8%; Pred. No. 64;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 RESULT 13
 AAY92504
 ID AAY92504 standard; Protein: 280 AA.
 AC AAY92504;
 DT 10-AUG-2000 (first entry)
 DE Human OXRE-1.
 XX
 XX OXRE-1; oxidoreductase; thioredoxin; antiproliferative; anticancer;
 KM hepatocytic; antiviral; antistimatic; nootropic; neuroprotectant;
 KM antiparkinsonian's; antisclerotic; anxiolytic; antischizophrenic;
 KM anti-inflammatory; antiarthritic.
 OS
 XX Homo sapiens.
 XX
 FH Key
 FT Domain
 FT 28..131 Location/Qualifiers
 FT /label= thioredoxin_domain
 FT 36..78 /note= "thioredoxin family active site"
 FT 68 /note= "potential phosphorylation site"
 FT 92 /note= "potential phosphorylation site"
 FT 119 /note= "potential phosphorylation site"
 FT 128 /note= "potential phosphorylation site"
 FT 128 /note= "potential phosphorylation site"
 FT 209 /note= "potential phosphorylation site"
 FT 221 /note= "potential phosphorylation site"
 FT 247 /note= "potential phosphorylation site"
 FT 257 /note= "potential phosphorylation site"
 FT 277 /note= "potential phosphorylation site"
 FT /note= "potential phosphorylation site"
 FT
 XX WO200020604-A2.
 PN
 XX
 PD 13-APR-2000.
 XX
 PE 06-OCT-1999; 99WO-US23434.
 XX
 PR 06-OCT-1998; 98US-0172227.
 PR 02-DEC-1998; 98US-0155202.
 PR 10-MAR-1999; 99US-0123911.
 XX

XX (INCY-) INCYTE PHARM INC.
PA
XX Lal P, Guegler KJ, Gorgone GA, Corley NC, Baughn MR, Tang YT;
PI Hillman JL, Bandman O, Azimzai Y, Au-Young J, Yue H, Lu DM;
PI Yang J;
XX
DR WPI: 2000-303785/26.
DR N-PSDB: AAA09375.
PT Purified polypeptide for treating or preventing disorders associated
PT with decreased expression or activity of oxidoreductase molecules
XX
PS Claim 1; Page 69-70; 97pp; English.
XX
CC AA92504-18 show OXRE-1 to -15. OXRE-1, an oxidoreductase, has identity
CC with thiorodoxin.
CC The polypeptides are useful for treating or preventing a disorder
CC associated with decreased expression or activity of OXRE. Antagonists
CC of OXRE are useful for treating or preventing a disorder associated with
CC increased expression or activity of OXRE. The disorders include cell
CC proliferative disorders (cirrhosis, hepatitis), cancer (leukemia,
CC melanoma), hypopituitarism and hyperpituitarism, hypothyroidism and
CC hyperthyroidism, metabolic disorders (Addison's disease, cystic
CC fibrosis), reproductive disorders (infertility, ovulatory defects),
CC neurological disorders (Alzheimer's disease, Parkinson's disease,
CC multiple sclerosis), mental disorders (anxiety, schizophrenia),
CC autoimmune/inflammatory disorders (acquired immunodeficiency syndrome
CC (AIDS), asthma, osteoarthritis), and viral infections. The
CC polynucleotides may be used in Southern or Northern analysis, polymerase
CC chain reaction (PCR), or in enzyme-linked immunosorbent assays (ELISA).
SQ Sequence 280 AA;

Query Match 50.6%; Score 44; DB 21; Length 280;
Best Local Similarity 53.8%; Pred. No. 64;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 MLVFFLLMGVTW 15
: | : ||||| |
DB 11 lavlvlllwgapw 23

RESULT 14

AA040021 standard; Protein: 280 AA.

AA040021;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 3166.

Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokine; thrombolytic; drug screening; arthritis; inflammation;
leukaemia.

Homo sapiens.

WO200153312-A1.

26-JUL-2001.

26-DEC-2000; 2000WO-US34263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
Zhao QH, Zhou P, Goodrich R, Drmanac RT;

WPI: 2001-442253/47.

N-PSDB: AA159177.

Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -

Example 4; SEQ ID NO 3166; 10078pp; English.

The invention relates to human nucleic acids (AA15798-AA161369) and
the encoded polypeptides (AA038642-AA042213) with noctropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilization of the activities such as: Immune system suppression,
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening
assays for receptor activity, arthritis and inflammation, leukaemias and
C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed
specification.

Sequence 280 AA;

Query Match 50.6%; Score 44; DB 22; Length 280;
Best Local Similarity 53.8%; Pred. No. 64;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 MLVFFLLMGVTW 15
: | : ||||| |
DB 11 lavlvlllwgapw 23

RESULT 15

AAB88357 standard; Protein: 280 AA.

AAB88357;

23-MAY-2001 (first entry)

Human membrane or secretory protein clone PSEC0085.

Human; secretory protein; membrane protein; vaccine; gene therapy;
rheumatoid arthritis; diabetes.

Homo sapiens.

EP1067182-A2.

10-JAN-2001.

07-JUL-2000; 2000EP-0114090.

08-JUL-1999; 99JP-0194179.

11-JAN-2000; 2000JP-0118775.

02-MAY-2000; 2000JP-0183766.

KW ischaemic shock; Alzheimer's disease; cognitive disorder; schizophrenia;
KW cardiovascular disorder; wound healing; stroke; arthritis; obesity;
KW asthma; sepsis; acne; psoriasis; transplant rejection; infection; AIDS;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
PN W09938881-A1.
XX
PD 05-AUG-1999.
XX
PF 27-JAN-1999; 99WO-US01621.
XX
PR 30-JAN-1998; 98US-0073170.
PR 30-JAN-1998; 98US-0073159.
PR 30-JAN-1998; 98US-0073160.
PR 30-JAN-1998; 98US-0073161.
PR 30-JAN-1998; 98US-0073162.
PR 30-JAN-1998; 98US-0073164.
PR 30-JAN-1998; 98US-0073165.
PR 30-JAN-1998; 98US-0073167.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Carter KC, Endress GA, Peng P, Ferrie AM, Florence C;
PI Florence KA, Janat F, Ni J, Rosen CA, Ruben SM;
PI Soppet DR, Young P, Yu G;
XX
DR WPI: 1999-469315/39.
DR N-PSDB: AA200455.
XX
PT New isolated human genes and the secreted polypeptides they encode
PT useful in, e.g. treatment of Alzheimer's
XX
PS Claim 1b; Page 323-324; 393pp; English.
XX
CC This invention describes novel human genes (see AA00410-200477) and the
CC secreted proteins (see AA025711-Y25778) and fragments (see
CC AA025779-Y25997) they encode. The polynucleotides and their corresponding
CC secreted polypeptides are useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. Also pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 67
CC polynucleotides of the invention, based on which tissues they are most
CC highly expressed in, and include developing products for the diagnosis
CC or treatment of cancer, tumours, neurodegenerative disorders,
CC developmental abnormalities and fetal deficiencies, blood disorders,
CC leukemias, diseases of the immune system, autoimmune diseases, hepatic
CC and renal disease, inflammation, allergies, ischaemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, restenosis, cardiovascular
CC disorders, wound healing, stroke, arthritis, obesity, asthma, sepsis,
CC acne, psoriasis, transplant rejection, metabolic disorders, infections
CC and AIDS. The polypeptides are also useful for identifying their binding
CC partners.
XX
SQ Sequence 281 AA:

Query Match 50.6%; Score 44; DB 20; Length 281;
Best Local Similarity 53.8%; Pred. No. 64;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 33314.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0121180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 50.6%; Score 44; DB 21; Length 312;
Best Local Similarity 60.0%; Pred. No. 72;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
OY 2 SMLVFLILMGVTWG 16
DB 283 sfllasflllgwtlig 297

RESULT 19
AAB23623
ID AAB23623 standard; Protein; 372 AA.
XX
AC AAB23623;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human secreted protein SEQ ID NO: 46.
XX
KW Human; secreted protein; cytokine; cell proliferation;
KW nutritional supplement; immune modulation; autoimmune disorder;

XX	haematopoiesis regulation; tissue growth; haemostasis; inflammation.
OS	Homo sapiens.
FH	Key
FT	Peptide
FT	Protein
XX	
PN	WO200049134-A1.
PB	
PD	24-AUG-2000.
XX	
PA	18-FEB-2000; 2000WO-US04340.
PR	19-FEB-1999; 99US-0120680.
PR	23-APR-1999; 99US-0298733.
PR	17-AUG-1999; 99US-0149639.
PR	23-SEP-1999; 99US-0155686.
PR	01-OCT-1999; 99US-0157247.
PR	29-NOV-1999; 99US-0167822.
PR	29-NOV-1999; 99US-0167823.
PR	15-FEB-2000; 2000US-0298733.
XX	
PA	(ALPH-) ALPHAGENE INC.
PI	Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
DR	WPI: 2000-549267/50.
XX	N-PDB: AAA93123.
PT	New secreted proteins and polynucleotides encoding them, which are derived from Homo sapiens, useful for therapy, diagnosis, and research, as well as nutritional sources or supplements -
PT	
PS	Claim 55; Page 272-274; 309pp; English.
XX	
CC	The present sequence is the sequence of a human secreted protein. Its cDNA was isolated from an adult lung cDNA library. The proteins
CC	and coding sequences of the invention can be used in the isolation of similar genes and proteins. In the elucidation of their function in vivo, and to treat a number of conditions. It is possible that they may have uses as nutritional supplements, as cytokine or cell proliferation factors, in immune modulation, where they may be used to treat immune and autoimmune diseases, as haematopoiesis regulators (treating myeloid or lymphoid cell deficiencies), in the promotion of tissue growth, they may have chemokine or chemotactic activity, haemostatic or thrombolytic activity, or anti-inflammatory activity.
SQ	Sequence 372 AA;
OY	Query Match 50.6%; Score 44; DB 21; Length 372; Best Local Similarity 53.3%; Pred. No. 87; Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0 2 SMLVFLLMGVTNG 16 : : Db 164 svltaflvllwlpnwg 178
RESULT 20	
ID AAB23643	
XX AAB23643 standard; Protein: 421 AA.	
AC AAB23643;	
DF 12-JAN-2001 (first entry)	
DE Human secreted protein SEQ ID NO: 99.	
KM Human secreted protein; cytokine; cell proliferation;	

KM nutritional supplement; immune modulation; autoimmune disorder;
 KM haematopoiesis regulation; tissue growth; haemostasis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WC0200049134-A1.
 XX
 PD 24-AUG-2000.
 XX
 PF 18-FEB-2000; 2000WO-US04340.
 XX
 PR 19-FEB-1999; 99US-0120680.
 PR 23-APR-1999; 99US-0298733.
 PR 17-AUG-1999; 99US-0149639.
 PR 23-SEP-1999; 99US-0155686.
 PR 01-OCT-1999; 99US-0157247.
 PR 29-NOV-1999; 99US-0167822.
 PR 29-NOV-1999; 99US-0167823.
 PR 15-FEB-2000; 2000US-0298733.
 XX
 PA (ALPH-) ALPHAGEN INC.
 XX
 PI Valenzuela D, Yuan O, Hoffman H, Hall J, Repiejko P;
 XX
 DR WPI: 2000-549267/50.
 XX
 PT New secreted proteins and polynucleotides encoding them, which are
 PT derived from Homo sapiens, useful for therapy, diagnosis, and research,
 PT as well as nutritional sources or supplements -
 XX
 PS Disclosure: Page 304-305; 309pp; English.
 XX
 CC The present invention is concerned with a number of secreted proteins
 CC and their coding sequences isolated from various human cDNA libraries.
 CC The proteins and coding sequences can be used in the isolation of
 CC similar genes and proteins. In the elucidation of their function in vivo,
 CC and to treat a number of conditions. It is possible that they may have
 CC uses as nutritional supplements, as cytokine or cell proliferation
 CC factors, in immune modulation, where they may be used to treat immune and
 CC autoimmune diseases, as haematopoiesis regulators (treating myeloid or
 CC lymphoid cell deficiencies), in the promotion of tissue growth, they may
 CC have chemokine or chemotactic activity, haemostatic or thrombolytic
 CC activity, or anti-inflammatory activity. No information about sequences
 CC AAB23632-B23645 is given in the specification.
 XX
 SO Sequence 421 AA:

Query Match 50.6%; Score 44; DB 21; Length 421;
 Best Local Similarity 53.3%; Pred. NO. 99;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0.

OY 2 SMLVYFLLMGVTWG 16
 I:I I:I I:I I:I
 Db 164 SVLTAFLVWIPHW 178

RESULT 21
 AAY57899
 ID AAY57899 standard; Protein: 439 AA.
 AC AAY57899;
 XX
 DT 23-MAR-2000 (first entry)
 XX
 DE Human transmembrane protein HTPMPN-23.
 XX
 KW Human: transmembrane protein; HTPMPN: diagnosis; immunospecific;
 KW antiproliferative; neuroprotective; immune disorder;
 KW reproductive disorder; smooth muscle disorder; neurological disorder;
 KW gastrointestinal disorder; developmental disorder;
 KW cell proliferative disorder.
 XX

OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 434
ET	/note= "unspecified"
XX	
PN	W09961471-A2.
XX	
PD	02-DEC-1999.
XX	
PE	28-MAY-1999; 99WO-US11904.
XX	
PR	29-MAY-1998; 98US-0087260.
PR	02-JUL-1998; 98US-0091674.
PR	02-OCT-1998; 98US-0102954.
PR	24-NOV-1998; 98US-0109869.
XX	
PA	(INCYTE) INCYTE PHARM INC.
XX	
PI	Tang YF, Lai P, Hillman JL, Yue H, Guegler KJ, Corley NC;
PI	Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;
PI	Au-Young J;
XX	
DR	WPI: 2000-072605/06.
DR	N-PDSB; AAZ56720.
XX	
PT	proteins, polynucleotides, vectors, host cells and antibodies used to
PT	diagnose, treat or prevent immune, reproductive, smooth muscle,
PT	neurological, gastrointestinal, developmental and cell proliferative
PT	disorders -
XX	
PS	Claim 1: Page 130-131; 229pp; English.
XX	
CC	AAZ56798 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC	transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively;
CC	The transmembrane protein have immunospecific, antiproliferative and
CC	neuroprotective activities. The human transmembrane proteins,
CC	polynucleotides encoding them and other compositions and methods from
CC	the present invention, can be used for the diagnosis, treatment or
CC	prevention of immune, reproductive, smooth muscle, neurological,
CC	gastrointestinal, developmental and cell proliferative disorders. The
CC	HTMPN's can be used to treat or prevent disorders associated with a
CC	decreased expression or activity of HTMPN.
XX	
SQ	Sequence 439 AA:
Query Match	50.6%; Score 44; DB 21; Length 439;
Best Local Similarity	53.3%; Pred. No. 1e+02;
Matches 8; Conservative	2; Mismatches 5; Indels 0; Gaps 0.
OY	2 SMLVFLLMGVTWC 16
I:I I I I I	
Dd	164 svltaflvlpwpg 178
RESULT 22	
ID	AAU14131 standard: Protein; 541 AA.
AAU14131	
AC	AAU14131;
DT	24-OCT-2001 (first entry)
DE	Human novel protein #2.
XX	
KW	Human; novel protein; antihaemic; osteopathic; antiinflammatory;
KW	immunomodulatory; cytosarctic; neuroprotective; vulnerary; nootropic;
KW	anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
KW	antibacterially; antibacterial; dermatologically; haemostatic; antischmatic;
KW	chromolytic; immunogen; antibody; gene therapy; neurological disorder;
KW	Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW	tissue regeneration; immune disorder.

XX	Homo sapiens.
OS	
XX	WO200155437-AZ.
PN	
XX	
PD	02-AUG-2001.
XX	
PF	25-JAN-2001; 2001WO-US02623.
XX	
PR	25-JAN-2000; 2000US-0491404.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YF, Liu C, Dymnac RT;
XX	
DR	WPI: 2001-451939/48.
N-PSDB:	AAS22436.
PT	
P7	Isolated polypeptides useful for treating anti-inflammatory diseases,
XX	nervous system disorders, and for regenerating bone and cartilage -
PS	Example 4: Page 520-521; 894pp: English.
XX	
CC	The invention relates to polynucleotides encoding novel human
CC	proteins or their active domains. The polypeptides, polynucleotides and
CC	antibodies raised against the polypeptides are used in a method of
CC	treatment of a mammal and prevention of disorders caused by the aberrant
CC	protein expression or activity. The polypeptides can be used as
CC	molecular weight markers, food supplements, and in antibody production.
CC	The polypeptides are used to identify compounds which bind to the
CC	polypeptides. Polynucleotides of the invention are used as probes and
CC	primers, for sequencing, for chromosome or gene mapping, in the
CC	production of recombinant proteins, and in generating anti-sense DNA or
CC	rRNA and in gene therapy. Polypeptides of the invention can be used to
CC	target drugs to a tumor, in assays to determine biological activity, to
CC	raise antibodies/elicit an immune response, to determine quantitative
CC	protein levels, as tissue markers, and to isolate receptors or ligands.
CC	Polypeptides of the invention may also be useful in treating platelet
CC	disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC	ligament and/or nerve tissue, wound healing, treating burns, promoting
CC	the proliferation, differentiation and survival of stem cells, as a
CC	contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC	Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC	sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC	fungal infection or from autoimmunity, cancer, allergy, asthma,
CC	grit-versus-host disease, eczema, haemophilia, thrombosis,
CC	anti-inflammatory diseases, nervous system disorders, and infection.
CC	The present sequence represents a protein of the invention.
XX	
SQ	Sequence 541 AA:
OY	2 SMLVVFLLMGVTWG 16
Db	I-I I-I I-I 164 svltarlvwiphwg 178
Query Match	50.6%; Score 44; DB 22; Length 541;
Best Local Similarity	53.3%; Pred. No. 1.3e+02;
Matches 8: Conservative	2; Mismatches 5; Indels 0; Gaps 0.
RESULT 23	
AAG89162	
ID AAG89162 standard: Protein; 541 AA.	
XX AAG89162;	
XX AC	
XX DT 11-SEP-2001 (first entry)	
XX DE Human secreted protein, SEQ ID NO: 282.	
XX KM Human: secreted protein; gene therapy; vaccine; treatment; diagnosis; GENSET.	

XX OS Homo sapiens.
XX PN WO200142451-A2.
XX PD 14-JUN-2001.
XX PE 07-DEC-2000; 2000WO-1B01938.
XX PR 08-DEC-1999; 99US-0169629.
XX PR 06-MAR-2000; 2000US-0187470.
XX PA (GEST) GENSET.
XX PI Dumas Milne Edwards J, Bougueleret L, Jobert S;
XX DR WPI: 2001-367870/38.
XX DR N-PSDB: AAH64765.
XX PT Full length GENSET human nucleic acids encoding potentially secreted
XX PT proteins, useful in gene therapy and vaccination against a variety of
XX PT diseases, and for diagnosis of those diseases -
XX PS Claim 21: Page 817-818; 921pp; English.
XX CC The invention relates to full length GENSET human nucleic acids encoding
XX CC potentially secreted proteins. The nucleic acids and the polypeptides
XX CC they encode may be used in the prevention, treatment and diagnosis of
XX CC diseases associated with inappropriate GENSET gene expression. For
XX CC example, they be used to treat disorders associated with decreased
XX CC GENSET gene expression by rectifying mutations or deletions in a
XX CC patient's genome that affect the activity of GENSET or by supplementing
XX CC the patient's own production of GENSET polypeptides. Conversely,
XX CC antisense nucleic acid molecules may be administered to down regulate
XX CC GENSET expression by binding with the cells' own genes and preventing
XX CC their expression. The sense and antisense nucleic acids may also be
XX CC used as DNA probes in diagnostic assays to detect and quantitate the
XX CC presence of similar nucleic acid sequences in samples, and hence to
XX CC determine which patients may be in need of restorative therapy.
XX CC The GENSET polypeptides may be used as antigens in the production of
XX CC antibodies and in assays to identify modulators (agonists and
XX CC antagonists) of GENSET polypeptide expression and activity. The
XX CC present sequence is a GENSET polypeptide of the invention.
XX SQ Sequence 541 AA:

Query Match 50.6%; Score 44; DB 22; Length 541;
Best Local Similarity 53.3%; Pred. No. 1.3e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 SMLVVFLLMGVTWG 16
1:1 11:11 11
Db 164 svltaflvlpw 178

RESULT 24
AAG89180
ID AAG89180 standard; Protein: 541 AA.
XX AC AAG89180;
XX DE 11-SEP-2001 (first entry)
XX DT Human secreted protein, SEQ ID NO: 300.
XX DE Human secreted protein, SEQ ID NO: 300.
XX KW Human; secreted protein; gene therapy; vaccine; treatment; diagnosis;
XX KM GENSET.
XX OS Homo sapiens.
XX OS WO200142451-A2.
XX PN XX

PD 14-JUN-2001.
XX PF 07-DEC-2000; 2000WO-1B01938.
XX PR 08-DEC-1999; 99US-0169629.
XX PR 06-MAR-2000; 2000US-0187470.
XX PA (GEST) GENSET.
XX PI Dumas Milne Edwards J, Bougueleret L, Jobert S;
XX DR WPI: 2001-367870/38.
XX DR N-PSDB: AAH64783.
XX PT Full length GENSET human nucleic acids encoding potentially secreted
XX PT proteins, useful in gene therapy and vaccination against a variety of
XX PT diseases, and for diagnosis of those diseases -
XX PS Claim 21: Page 829-831; 921pp; English.
XX CC The invention relates to full length GENSET human nucleic acids encoding
XX CC potentially secreted proteins. The nucleic acids and the polypeptides
XX CC they encode may be used in the prevention, treatment and diagnosis of
XX CC diseases associated with inappropriate GENSET gene expression. For
XX CC example, they be used to treat disorders associated with decreased
XX CC GENSET gene expression by rectifying mutations or deletions in a
XX CC patient's genome that affect the activity of GENSET or by supplementing
XX CC the patient's own production of GENSET polypeptides. Conversely,
XX CC antisense nucleic acid molecules may be administered to down regulate
XX CC GENSET expression by binding with the cells' own genes and preventing
XX CC their expression. The sense and antisense nucleic acids may also be
XX CC used as DNA probes in diagnostic assays to detect and quantitate the
XX CC presence of similar nucleic acid sequences in samples, and hence to
XX CC determine which patients may be in need of restorative therapy.
XX CC The GENSET polypeptides may be used as antigens in the production of
XX CC antibodies and in assays to identify modulators (agonists and
XX CC antagonists) of GENSET polypeptide expression and activity. The
XX CC present sequence is a GENSET polypeptide of the invention.
XX SQ Sequence 541 AA:

Query Match 50.6%; Score 44; DB 22; Length 541;
Best Local Similarity 53.3%; Pred. No. 1.3e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 SMLVVFLLMGVTWG 16
1:1 11:11 11
Db 164 svltaflvlpw 178

RESULT 25
AAV01453
ID AAV01453 standard; Protein: 40 AA.
XX AC AAV01453;
XX DE 18-MAY-1999 (first entry)
XX DT Secreted protein encoded by gene 63 clone HJAT30.
XX DE Secreted protein encoded by gene 63 clone HJAT30.
XX KW Human; secreted protein; gene therapy; protein therapy; cancer; weight;
XX KM tumour; chromosome mapping; forensic; hematological disease; allergy;
XX KM inflammation; cell proliferation; viral infection; wound healing;
XX KM modulation; appetite; behaviour; food additive; preservative.
XX OS Homo sapiens.
XX OS WO9903990-A1.
XX PN 28-JAN-1999.
XX PD 15-JUL-1998; 98WO-US14613.
XX PF

```
XX 18-AUG-1997; 97US-0056361.
PR 16-JUL-1997; 97US-0052661.
PR 16-JUL-1997; 97US-0052870.
PR 16-JUL-1997; 97US-0052871.
PR 16-JUL-1997; 97US-0052872.
PR 16-JUL-1997; 97US-0052873.
PR 16-JUL-1997; 97US-0052874.
PR 16-JUL-1997; 97US-0052875.
PR 22-JUL-1997; 97US-0053440.
PR 22-JUL-1997; 97US-0053441.
PR 22-JUL-1997; 97US-0053442.
PR 18-AUG-1997; 97US-0053683.
PR 18-AUG-1997; 97US-0055724.
PR 18-AUG-1997; 97US-0055725.
PR 18-AUG-1997; 97US-0055726.
PR 18-AUG-1997; 97US-0055946.
PR 18-AUG-1997; 97US-0055952.
PR 18-AUG-1997; 97US-0055985.
PR 18-AUG-1997; 97US-0055989.
PR 18-AUG-1997; 97US-0056359.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Duan R, Feng P, Ferric AM, Florence KA, Fouda J;
PI Greene JM, Hu J, NI J, Rosen CA, Ruben SM, Young PE;
PI Yu G;
XX
XX WPI: 1999-132234/11.
DR N-PSDB: AAX22281.
XX
XX New nucleic acids encoding secreted human proteins - potentially
PT useful for treating and diagnosing diseases and identifying specific
PT binding agents
XX
XX Claim 11; Page 238; 251pp; English.
XX
XX The invention relates to nucleic acid sequences (AAX22211 to AAX22282)
CC encoding human secreted proteins (AAV01383 to AAV01454). The secreted
CC protein gene sequences are deposited with the ATCC under deposit number
CC ATCC 209138, 209139 or 209141. Host cells containing vectors comprising
CC the nucleic acid sequences are used for the recombinant expression of
CC the secreted proteins. The polynucleotide and amino acid sequences are
CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein or gene therapy. Pathological conditions can be also diagnosed
CC by determining the amount of the new polypeptides in a sample or by the
CC presence of mutations in the new polynucleotides. The nucleic acid
CC sequences, or its fragments, are useful for chromosome identification
CC and mapping; as antisense and triplex-forming therapeutics; in gene
CC therapy; for (forensic) identification of individuals; as molecular
CC weight markers; to identify related sequences or specific mRNA; in
CC preparation of oligomers and to raise anti-DNA antibodies. Antibodies are
CC useful as immunosay reagents (including for in vivo imaging) and
CC therapeutically to inhibit or activate particular polypeptides. A very
CC wide range of disorders may be treated with the polynucleotide and
CC polypeptide sequences, e.g. autoimmune or haematological diseases,
CC allergy, inflammation, cancer or other forms of cell proliferation, viral
CC or other infections. The sequences may also be useful in wound healing,
CC to modulate differentiation of embryonic stem cells, to modulate weight,
CC appetite, behaviour etc. and as food additive or preservative. The
CC present sequence represents a human secreted protein (see descriptor
CC line for gene number and clone identification).
XX
XX Sequence 40 AA:
SQ
```

```
RESULT 26
AAU03680
ID AAU03680 standard; Protein: 65 AA.
XX
XX AAU03680;
AC
XX
XX 12-SEP-2001 (first entry)
DT
XX
XX Group B Streptococcus antigenic protein, ID-157.
DE
XX
XX Group B Streptococcus; encapsulated bacterium; therapeutic; sepsis;
XX meningitis; neonate; antigenic; vaccine; infection; genital tract;
XX capsid polysaccharide vaccination.
OS
XX Streptococcus agalactiae.
XX
XX WO200132882-A2.
XX
XX 10-MAY-2001.
PD
XX
XX 07-SEP-2000; 2000MO-GB03437.
PE
XX
XX 07-SEP-1999; 99GB-0021125.
PR
XX
XX (MICR-) MICROBIAL TECHNIQS LTD.
PA
XX
XX Le Page RWF, Wells JM, Hanniffy SB;
PI
XX
XX WPI: 2001-316444/33.
DR
XX
XX N-PSDB: AAS07097.
XX
XX New polypeptides derived from Streptococcus agalactiae are useful to
PT provide detection of, and vaccination against, Group B Streptococcus
PT infections, particularly to prevent infection in neonates -
XX
XX Claim 1; Fig 1; 178pp; English.
XX
XX AAU03601-AAU03722 represent Group B Streptococcus (Streptococcus
CC agalactiae) amino acid sequences of the invention. S. agalactiae is an
CC encapsulated bacterium which is a major pathogen of humans causing sepsis
CC and meningitis in neonates as well as adults. The S. agalactiae antigenic
CC polypeptides are used to vaccinate against Group B Streptococcus
CC infections, particularly to prevent infection in new born children
CC arising from the maternal genital tract. An immunogenic composition is
CC useful in the preparation of a medicament for the treatment or
CC prophylaxis of Group B Streptococcus infection. The invention does not
CC have the disadvantages of varied response rate associated with prior art
CC capsid polysaccharide vaccination against Group B Streptococcus.
XX
XX Sequence 65 AA:
SQ
```

```
Query Match 49.4%; Score 43; DB 22; Length 65;
Best Local Similarity 45.5%; Pred. No. 19;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
OY 6 VFLLMGVTVWG 16
:1111111111
DB 15 IFLIWIWIIWG 25

RESULT 27
AAR80565
ID AAR80565 standard; Protein: 137 AA.
XX
XX AAR80565;
AC
XX
XX 17-JAN-1996 (first entry)
DT
XX
XX EAV GL protein.
DE
XX
```

KW GL protein; diagnostic; vaccine; equine viral arteritis.
 XX Equine arteritis virus.
 OS
 XX WO9519438-A1.
 XX 20-JUL-1995.
 PD
 XX 13-JAN-1995; 95WO-GB00066.
 PR
 XX 14-JAN-1994; 94GB-0000656.
 PR
 XX (UKAG-) UK MIN FISHERIES & FOOD.
 PA
 XX Chlrmide ED:
 PI
 XX WPI; 1995-263868/34.
 DR
 XX N-PSDB; AA098814.
 XX
 XX New equine arteritis virus peptide(s), conjugates, DNA and
 PT antibodies are used as vaccines or diagnostic agents in equine
 PT arteritis virus-mediated diseases
 PS
 PS Claim 1: Page 18; 31pp; English.
 XX
 CC The DNA sequence equivalent to the entire EAV genome minus the
 CC first 18 bases and the poly-A tail is given in AA098814, and
 CC includes the GL protein coding sequence. Epitopes of GL are
 CC used in vaccine prodn.
 CC
 XX
 SQ Sequence 137 AA;

Query Match 49.4%; Score 43; DB 16; Length 137;
 Best Local Similarity 58.3%; Pred. No. 42;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSMLVPLLMG 12
 :||:||||
 Db 2 lsmivllflwg 13

RESULT 28
 AA74573
 ID AA74573 standard; Protein: 218 AA.
 XX
 AC AA74573;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria gonorrhoeae ORF 152 protein sequence SEQ ID NO:620.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX
 OS Neisseria gonorrhoeae.
 OS
 XX WO9957280-A2.
 PN
 XX 11-NOV-1999.
 PD
 XX 30-APR-1999; 99WO-US09346.
 PF
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX

PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Piza M, Rappuoli R, Ratli G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 XX WPI; 2000-062150/05.
 DR
 XX N-PSDB; AA253335.
 DR
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 PT
 XX
 XX Claim 2: Page 435; 1453pp; English.
 PS
 XX AA253015 to AA254536, AA254577 to AA254615, and AA74253 to AA75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 CC
 XX
 SQ Sequence 218 AA;

Query Match 49.4%; Score 43; DB 21; Length 218;
 Best Local Similarity 43.8%; Pred. No. 68;
 Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSMLVPLLMGVTWG 16
 :||:||||:|
 Db 43 vglivllivrlcwg 58

RESULT 29
 AA74574
 ID AA74574 standard; Protein: 218 AA.
 XX
 AC AA74574;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 152 protein sequence SEQ ID NO:622.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX
 OS Neisseria meningitidis.
 OS
 XX WO9957280-A2.
 PN
 XX 11-NOV-1999.
 PD
 XX 30-APR-1999; 99WO-US09346.
 PF
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
DR N-PSDB: AA253336.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2: Page 435; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC *Neisserial* bacteria (e.g. meningitis and septicemia), to detect the
CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
XX Sequence 218 AA:
SQ

Query Match 48.9%; Score 42.5; DB 21; Length 218;
Best Local Similarity 61.5%; Pred. No. 80;
Matches 8; Conservative 2; Mismatches 2; Indels 1; Gaps 1;
OY 4 LVVFLLMKCVTWG 16
| : | | | : | : | : |
Db 50 llvflrcwgl-wg 61

RESUL.T 30
AAY74575
ID AAY74575 standard; Protein: 218 AA.
XX
XX AAY74575;
AC
XX
XX 21-MAR-2000 (first entry)
DT
XX
XX *Neisseria meningitidis* ORF 152 protein sequence SEQ ID NO:624.
DE
XX
XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
XX *Neisseria meningitidis*.
OS
XX
XX WO9957280-A2.
PN
XX
XX 11-NOV-1999.
PD
XX
XX 30-APR-1999; 99WO-US09346.
PF
XX
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.

XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
DR N-PSDB: AA253337.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2: Page 436; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC *Neisserial* bacteria (e.g. meningitis and septicemia), to detect the
CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
XX Sequence 218 AA:
SQ

Query Match 48.9%; Score 42.5; DB 21; Length 218;
Best Local Similarity 61.5%; Pred. No. 80;
Matches 8; Conservative 2; Mismatches 2; Indels 1; Gaps 1;
OY 4 LVVFLLMKCVTWG 16
| : | | | : | : | : |
Db 50 llvflrcwgl-wg 61

RESUL.T 31
AAB64795
ID AAB64795 standard; Protein: 37 AA.
XX
XX AAB64795;
AC
XX
XX 23-MAR-2001 (first entry)
DT
XX
XX Human secreted protein sequence encoded by gene 23 SEQ ID NO:81.
DE
XX
XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; autoimmune disease; hyperproliferative disorder;
KW cardiovascular disorder; cerebrovascular disorder; infection; chemotaxis;
KW nervous system disorder; ocular disorder; skin aging; wound healing;
KW food additive; tissue regeneration.
XX
XX
XX Homo sapiens.
OS
XX
XX WO200077256-A1.
PN
XX
XX 21-DEC-2000.
PD
XX
XX 01-JUN-2000; 2000WO-US14963.
PF
XX
XX 11-JUN-1999; 99US-0138631.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM, Komatsoulis GA;
PI
XX
XX WPI: 2001-032315/04.
DR
XX
XX N-PSDB: AAF33117.

XX Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 XX
 PS Claim 11; Page 464; 506pp; English.
 XX
 CC Polynucleotide sequences AAF33095 - AAF31142 encode human secreted
 CC proteins AAB64773 - AAB64820. Fragments of the secreted proteins and
 CC amino acid sequences which share homology with the fragments are
 CC represented in AAB64821 - AAB64880. The genes and proteins have
 CC activities dependent on the tissues and cells in which they are
 CC expressed. Examples of their activities and the activities of their
 CC agonists and antagonists include: immunosuppressive; antirheumatic;
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 CC cerebroprotective; neuroprotective; antibacterial; virucide;
 CC fungicide; and ophthalmological. The secreted proteins, polynucleotides,
 CC antagonists and agonists may be useful in treating, preventing and
 CC diagnosing diseases and disorders such as autoimmune diseases e.g.
 CC rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the
 CC breast or liver, cardiovascular disorders e.g. cardiac arrest,
 CC cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous
 CC system disorders e.g. Alzheimer's disease, infections caused by bacteria,
 CC viruses and fungi and ocular disorders e.g. corneal infection. The
 CC polypeptides can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues and in chemotaxis. The polypeptides can also be
 CC used as a food additive or preservative to increase or decrease storage
 CC capabilities. Included in the invention are sequences AAB64772 and
 CC AAF33095 - AAF33142 which are used in the isolation and characterisation
 CC of the nucleotide and protein sequences of the invention.

XX Sequence 37 AA:

Query Match 48.3%; Score 42; DB 22; Length 37;

Best Local Similarity 63.6%; Pred. No. 15;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 SMLVFLLMG 12

Db 4 SLIAFLLIwG 14

RESULT 32

AA676763 AAB76763 standard; Protein; 402 AA.

XX AAB76763;

XX 11-APR-2001 (first entry)

DE Corynebacterium glutamicum MCT protein SEQ ID NO:508.

XX Corynebacterium glutamicum: brevbacterium lactofermentum; MCT;

KM membrane construction and membrane transport protein; petroleum spill;

KW hydrocarbon degradation; gram positive aerobic bacterium; marker;

KM identification; microorganism; fine chemical production; transformation;

OS genome mapping; genetic engineering.

XX Corynebacterium glutamicum.

XX WO200100805-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000MO-IB00926.

XX 25-JUN-1999; 99US-0141031.

XX 08-JUL-1999; 99DE-1031454.

XX 08-JUL-1999; 99DE-1031478.

XX 08-JUL-1999; 99DE-1031563.

XX 09-JUL-1999; 99DE-1032122.

PR 09-JUL-1999; 99DE-1032124.

PR 09-JUL-1999; 99DE-1032125.

PR 09-JUL-1999; 99DE-1032126.

PR 09-JUL-1999; 99DE-1032127.

PR 09-JUL-1999; 99DE-1032128.

PR 09-JUL-1999; 99DE-1032129.

PR 09-JUL-1999; 99DE-1032130.

PR 09-JUL-1999; 99DE-1032131.

PR 09-JUL-1999; 99DE-1032132.

PR 09-JUL-1999; 99DE-1032133.

PR 09-JUL-1999; 99DE-1032134.

PR 09-JUL-1999; 99DE-1032135.

PR 09-JUL-1999; 99DE-1032136.

PR 09-JUL-1999; 99DE-1032137.

PR 09-JUL-1999; 99DE-1032138.

PR 09-JUL-1999; 99DE-1032139.

PR 09-JUL-1999; 99DE-1032140.

PR 09-JUL-1999; 99DE-1032141.

PR 09-JUL-1999; 99DE-1032142.

PR 09-JUL-1999; 99DE-1032143.

PR 09-JUL-1999; 99DE-1032144.

PR 09-JUL-1999; 99DE-1032145.

PR 09-JUL-1999; 99DE-1032146.

PR 09-JUL-1999; 99DE-1032147.

PR 09-JUL-1999; 99DE-1032148.

PR 09-JUL-1999; 99DE-1042078.

PR 09-JUL-1999; 99DE-1042079.

PR 09-JUL-1999; 99DE-1042080.

PR 09-JUL-1999; 99DE-1042081.

PR 09-JUL-1999; 99DE-1042082.

PR 09-JUL-1999; 99DE-1042083.

PR 09-JUL-1999; 99DE-1042084.

PR 09-JUL-1999; 99DE-1042085.

PR 09-JUL-1999; 99DE-1042086.

PR 09-JUL-1999; 99DE-1042087.

PR 09-JUL-1999; 99DE-1042088.

PR 09-JUL-1999; 99DE-1042089.

PR 09-JUL-1999; 99DE-1042090.

PR 09-JUL-1999; 99DE-1042091.

PR 09-JUL-1999; 99DE-1042092.

PR 09-JUL-1999; 99DE-1042093.

PR 09-JUL-1999; 99DE-1042094.

PR 09-JUL-1999; 99DE-1042095.

PR 09-JUL-1999; 99DE-1042096.

PR 09-JUL-1999; 99DE-1042097.

PR 09-JUL-1999; 99DE-1042098.

PR 09-JUL-1999; 99DE-1042099.

PR 09-JUL-1999; 99DE-1042100.

PR 09-JUL-1999; 99DE-1042101.

PR 09-JUL-1999; 99DE-1042102.

PR 09-JUL-1999; 99DE-1042103.

PR 09-JUL-1999; 99DE-1042104.

PR 09-JUL-1999; 99DE-1042105.

PR 09-JUL-1999; 99DE-1042106.

PR 09-JUL-1999; 99DE-1042107.

PR 09-JUL-1999; 99DE-1042108.

PR 09-JUL-1999; 99DE-1042109.

PR 09-JUL-1999; 99DE-1042110.

PR 09-JUL-1999; 99DE-1042111.

PR 09-JUL-1999; 99DE-1042112.

PR 09-JUL-1999; 99DE-1042113.

PR 09-JUL-1999; 99DE-1042114.

PR 09-JUL-1999; 99DE-1042115.

PR 09-JUL-1999; 99DE-1042116.

PR 09-JUL-1999; 99DE-1042117.

PR 09-JUL-1999; 99DE-1042118.

PR 09-JUL-1999; 99DE-1042119.

PR 09-JUL-1999; 99DE-1042120.

PR 09-JUL-1999; 99DE-1042121.

PR 09-JUL-1999; 99DE-1042122.

PR 09-JUL-1999; 99DE-1042123.

PR 09-JUL-1999; 99DE-1042124.

PR 09-JUL-1999; 99DE-1042125.

PR 09-JUL-1999; 99DE-1042126.

PR 09-JUL-1999; 99DE-1042127.

```

XX 26-SEP-2001 (first entry)
DT C glutamicum protein fragment SEQ ID NO: 6555.
XX
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX Corynebacterium glutamicum.
XX
XX EPI108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX
XX 07-APR-2000; 2000JP-0159162.
XX
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
XX
XX N-PSDB; AAH68020.
XX
XX Novel polynucleotides derived from Corynebacterio, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 17; SEQ ID NO: 6555; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Corynebacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of corynebacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Corynebacterium, and identifying a homologue of a gene derived
XX from Corynebacterium. Corynebacterium bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a protein described
XX in the exemplification of the invention.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.
XX
XX Sequence 449 AA:
SQ

```

```

Query Match 48.3%; Score 42; DB 22; Length 449;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
OY 4 LVVFLLMGVTW 15
DB 424 LVLFGLAWGLKW 435

```

```

RESULT 34
AARI2362
ID AARI2362 standard; Protein; 455 AA.
XX
XX AARI2362;
XX
XX 13-SEP-1991 (first entry)
XX
XX Octopus rhodopsin membrane protein.
XX
XX Octopus; rhodopsin; membrane; helix; OR.
XX
XX

```

```

PN JP03123486-A.
XX
XX 27-MAY-1991.
XX
XX 06-OCT-1989; 89JP-0260261.
XX
XX 06-OCT-1989; 89JP-0260261.
XX
XX (HITA ) HITACHI KK.
XX
XX WPI; 1991-197925/27.
XX
XX N-PSDB; AA012225.
XX
XX Genetic engineering of membrane protein - by division of protein
XX into cartridge genes corresp. to helix structure polypeptides)
XX
XX Disclosure; Fig 1; 17pp; Japanese.
XX
XX A gene cassette is prepd. by division of the OR membrane protein
XX gene encoding helix structure polypeptides. A base sequence contg.
XX such a sequence downstream to the tryptophan regulating gene derived
XX from the E. coli tryptophan operon, the trpL, or the trpE polypeptide
XX translation regulating base sequence, and the N-terminal Met of the trpL
XX (or trpE) or trpE polypeptide, respectively, is introduced into
XX an expression vector for transformation of host cells. A partial
XX or total OR membrane protein is produced by culturing the
XX transformants.
XX
XX Sequence 455 AA:
SQ

```

```

Query Match 48.3%; Score 42; DB 12; Length 455;
Best Local Similarity 23.1%; Pred. No. 2.1e+02;
Matches 3; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
OY 3 MLVFLLMGVTW 15
DB 155 LMILFVWMSLW 167

```

```

RESULT 35
AA09518
ID AA09518 standard; Protein; 596 AA.
XX
XX AA09518;
XX
XX 16-JUL-1999 (first entry)
XX
XX C. elegans dopamine transport protein CedAT1.
XX
XX C. elegans; dopamine transport protein; CedAT2; CedAT1; nematode;
XX antiparasitic; psychoactive drug.
XX
XX Caenorhabditis elegans.
XX
XX WO9921883-A1.
XX
XX 06-MAY-1999.
XX
XX 27-OCT-1998; 98WO-US22712.
XX
XX 27-OCT-1997; 97US-0063282.
XX
XX (AMCY ) AMERICAN CYANAMID CO.
XX (UYVA-) UNIV VANDERBILT.
XX
XX Blakely RD, Epler CM;
XX
XX WPI; 1999-312943/26.
XX
XX N-PSDB; AAX56244.
XX
XX Novel C.elegans dopamine transporter proteins
XX
XX

```

PS Claim 16; Page 35-37; 47pp; English.

XX The present sequence represents the *C. elegans* (nematode) dopamine transporter *CedAT1*. *CedAT1* is a shortened version of *CedAT2*. The present

CC invention also describes: (1) an expression vector comprising a DNA sequence operably linked to a control sequence which is compatible with

CC a suitable host cell; (2) an expression system comprising a host cell transformed with the expression vector of (1); (3) a method of producing

CC purified dopamine receptor; (4) an antibody to a dopamine receptor; (5) a DNA probe comprising a nucleic acid as defined above. The expression

CC system is used in a screening assay for dopamine transporter antagonists or stimulants. These can be used as antiparasitic or psychoactive drugs, especially antihelminthic drugs. The nucleic acid can be used to

CC determine the presence of other dopamine transporter genes in an organism. The probe may be used for detecting genes encoding dopamine transporters.

CC

SQ Sequence 596 AA:

Query Match 48.3%; Score 42; DB 20; Length 596;
Best Local Similarity 35.7%; Pred. No. 2.7e+02;
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

OY 3 MLVFFLLMGVTWG 16
:::|:::|111
Db 444 lImewlllygtlwg 457

RESULT 36
AAY09517
ID AAY09517 standard; Protein: 615 AA.

XX AAY09517;
AC AAY09517;
XX 16-JUL-1999 (first entry)
DT
XX
XX
DE C. elegans dopamine transport protein *CedAT2*.
XX
XX C. elegans; dopamine transport protein; *CedAT2*; *CedAT1*; nematode;
KM antiparasitic; psychoactive drug.
XX
XX *Caenorhabditis elegans*.
OS
XX
XX WO921883-A1.
PN
XX 06-MAY-1999.
PD
XX
XX 27-OCT-1998; 98WO-US22712.
PF
XX 27-OCT-1997; 97US-0063282.
PR
XX 27-OCT-1997; 97US-0063282.
XX
XX (AMCY) AMERICAN CYANAMID CO.
PA (UYVA-) UNIV VANDERBILT.
XX
XX Blakely RD, Eppler CM;
PI
XX
XX WPT: 1999-312943/26.
DR N-PSDB: AAX56244.
DR
XX
XX Novel C.elegans dopamine transporter proteins
PT
XX
XX Claim 15; Page 33-35; 47pp; English.

XX The present sequence represents the *C. elegans* (nematode) dopamine transporter *CedAT2*. *CedAT1* is a shortened version of *CedAT2*. The present

CC invention also describes: (1) an expression vector comprising a DNA sequence operably linked to a control sequence which is compatible with

CC a suitable host cell; (2) an expression system comprising a host cell transformed with the expression vector of (1); (3) a method of producing

CC purified dopamine receptor; (4) an antibody to a dopamine receptor; (5) a DNA probe comprising a nucleic acid as defined above. The expression

CC system is used in a screening assay for dopamine transporter antagonists

CC or stimulators. These can be used as antiparasitic or psychoactive drugs, especially antihelminthic drugs. The nucleic acid can be used to

CC determine the presence of other dopamine transporter genes in an organism. The probe may be used for detecting genes encoding dopamine transporters.

CC

SQ Sequence 615 AA:

Query Match 48.3%; Score 42; DB 20; Length 615;
Best Local Similarity 35.7%; Pred. No. 2.8e+02;
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

OY 3 MLVFFLLMGVTWG 16
:::|:::|111
Db 463 lImewlllygtlwg 476

RESULT 37
AAB39016
ID AAB39016 standard; Protein: 65 AA.

XX AAB39016;
AC AAB39016;
XX
XX 02-FEB-2001 (first entry)
DT
XX
XX
DE Human secreted protein #14.
XX
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KM vulnerability; anticonvulsant; antibacterial; antifungal; antiparasitic;
KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KM neurological disease; infection; human; secreted protein.
XX
XX
XX *Homo sapiens*.
OS
XX
XX WO200056880-A1.
PN
XX
XX 28-SEP-2000.
PD
XX
XX 16-MAR-2000; 2000WO-US06781.
PF
XX
XX 19-MAR-1999; 99US-0125363.
PR 08-DEC-1999; 99US-0169617.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
PI
XX
XX WPT: 2000-602220/57.
DR N-PSDB: AAC59724.
DR
XX
XX Nucleic acid molecules encoding human secreted proteins, used in

PT preventing, treating or ameliorating disorders such as Parkinson's and

PT Alzheimer's diseases, cancers and infections -

XX
XX
XX Claim 11; Page 387; 422pp; English.

XX Sequences AAB38971-839020 represent the amino acid sequences of 50

CC human secreted proteins encoded by the genes AAC59679-C59728. The genes

CC and proteins are useful for preventing, ameliorating or treating medical

CC conditions, e.g. by protein or gene therapy. The genes are isolated from

CC a range of human tissues disclosed in the specification. The nucleic

CC acids, proteins, antibodies and (ant)agonists are useful in the

CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and

CC ovarian cancer, and other cancers of the adrenal gland, bone, bone

CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;

CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune

CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,

CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative

CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;

CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and

CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

CC and parasitic infections.

XX Sequence 65 AA:

Query Match 47.7% Score 41.5; DB 21; Length 65;
Best Local Similarity 56.2%; Pred. No. 31;
Matches 9; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

OY 1 MSMLVFL-LMGVTW 15
| | : : | | | | |
Db 1 mrlalwllglwqvmw 16

RESULT 38

AAM25231

ID AAM25231 standard; Protein: 135 AA.

XX AC AAM25231;

XX DT 16-OCT-2001 (first entry)

DE Human protein sequence SEQ ID NO:746.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;
KW antiagregant; haemostatic; vulnery; antilucer; osteopathic; eczema;
KW dermatologic; antiallergic; antisthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidiopressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.

XX OS Homo sapiens.

XX PN WO200153455-A2.

XX PD 26-JUL-2001.

XX PF 22-DEC-2000; 2000MO-US35017.

XX PR 23-DEC-1999; 990S-0471275.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI: 2001-457603/49.

XX DR N-PSDB; AAH99172.

PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
XX
PS Claim 20; Page 181; 1217pp; English.

CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAH55963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianemic; antiagregant; haemostatic; vulnery;
CC antilucer; osteopathic; dermatologic; antiallergic; antisthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidiopressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides

CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.

XX Sequence 135 AA:

Query Match 47.7% Score 41.5; DB 22; Length 135;
Best Local Similarity 69.2%; Pred. No. 67;
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

OY 4 LVVFLLMGVTWG 16
| | : : | | | | |

Db 122 lvflfllwsv-wg 133

RESULT 39

AAH82774

ID AAH82774 standard; Protein: 227 AA.

XX AC AAH82774;

XX DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:2642.

XX Staphylococcus epidermidis SRI strain; infection; diagnosis;

XX KW vaccination; endocarditis.

XX OS Staphylococcus epidermidis.

XX PN WO200134809-A2.

XX PD 17-MAY-2001.

XX PF 09-NOV-2000; 2000MO-US30782.

XX PR 09-NOV-1999; 990S-0164258.

XX (GLAX) GLAXO GROUP LTD.

XX PI Kimerly MJ;

XX DR WPI: 2001-316495/33.

XX DR N-PSDB; AAH53624.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX
PS Claim 18; Page 693; 2188pp; English.

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAH81454 to AAH83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to

CC AAH5098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 227 AA:

Query Match 47.7%; Score 41.5; DB 22: Length 227;
Best Local Similarity 61.5%; Pred. No. 1.2e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

OY 5 VVFLILMGVTVWG 16
: : | | | | | | | | | |
Db 125 IVLILMGVTVWG 137

RESULT 40
AAU01020
ID AAU01020 standard: Protein: 304 AA.
XX
AC AAU01020:
XX
DT 04-JUL-2001 (first entry)
XX
DE Human secreted soluble alpha2delta calcium channel subunit #7 protein.
XX
KW Human: secreted calcium channel alpha2delta subunit: alpha2delta-2;
KW alpha2delta-3; alpha2delta-4; voltage-dependent calcium channel; VDCC;
KW gabapentin; scintillation proximity assay; SPA; nickel flashplate assay;
KW filter binding assay; wheat germ lectin flashplate assay.
XX
OS Homo sapiens.
XX
PM WO200119870-A2.
XX
PD 22-MAR-2001.
XX
PF 18-SEP-2000; 2000WO-EP09137.
XX
PR 16-SEP-1999; 99US-0397550.
XX
PA (WARN) WARNER LAMBERT CO.
XX
PI Brown JP, Bertelli F;
XX
DR WPI: 2001-235262/24.
DR N-PSDB: AAS01410.
XX
PT Calcium channel alpha2delta subunits, useful in e.g. SPA assays,
PT Flashplate assays, Nickel Flashplate assays, Filter binding assays or
PT Wheat Germ Lectin Flashplate assays -
XX
PS Disclosure: Page 84: 160pp: English.
XX
CC The present sequence represents human secreted calcium channel
CC alpha2delta subunit #7 which is soluble and retains the functional
CC characteristics of the full length or wild type alpha2delta subunit
CC (AAU01025) from which it is derived. The invention relates to truncated
CC alpha2delta-2, alpha2delta-3 or alpha2delta-4 subunit soluble proteins
CC which retain their affinity for radioactively labelled gabapentin. The
CC alpha2delta subunit is 1 of the components of the heteromultimeric
CC voltage-dependent calcium channel (VDCC) complexes present in neuronal
CC and non-neuronal tissues including heart and skeletal muscle. Numerous
CC soluble forms of the human calcium channel alpha2delta subunits
CC (AAU01014-AAU01024 and AAU01032-AAU01038) and 5 soluble forms of the
CC porcine calcium channel alpha2delta subunits (AAU01027-AAU01031) are
CC described. The secreted soluble alpha2delta subunit may be used in assays
CC e.g. scintillation proximity assay (SPA), flashplate, nickel flashplate,
CC filter binding or wheat germ lectin flashplate assays to detect or

CC measure the binding or interaction of a ligand (e.g. gabapentin,
CC L-Norleucine, L-Allo-Isoleucine, L-methionine, L-leucine, L-Isoleucine,
CC L-Valine, Spermine and/or L-Phenylalanine) of a calcium channel
CC alpha2delta subunit.
XX
SQ Sequence 304 AA:

Query Match 47.7%; Score 41.5; DB 22: Length 304;
Best Local Similarity 69.2%; Pred. No. 1.6e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

OY 4 LVVFLILMGVTVWG 16
: : | | | | | | | | | |
Db 209 IVLILMGVTVWG 220

RESULT 41
AAB62244
ID AAB62244 standard: Protein: 304 AA.
XX
AC AAB62244:
XX
DT 11-JUN-2001 (first entry)
XX
DE Human calcium channel alpha2delta subunit related seq ID No. 16.
XX
KW Calcium channel alpha2delta subunit: alpha2delta-1; cerebral cortex;
KW nervous system disorder; pain; epilepsy; anxiety; human.
XX
OS Homo sapiens.
XX
PM WO200120336-A2.
XX
PD 22-MAR-2001.
XX
PF 18-SEP-2000; 2000WO-EP09136.
XX
PR 16-SEP-1999; 99US-0397549.
XX
PA (WARN) WARNER LAMBERT CO.
XX
PI Bertelli F, Brown JP, Dissanayake V, Suman-Chauhan N, Gee NS;
XX
DR WPI: 2001-257902/26.
DR N-PSDB: AAF57547.
XX
PT Competitive binding assay for screening ligands which bind a cerebral
PT cortical voltage-dependent calcium channel alpha2delta-1 subunit,
PT where the ligands identified are useful for treating disorders of the
PT nervous system, including pain -
XX
PS Disclosure: Page 84: 158pp: English.
XX
CC The invention relates to a new method for screening ligands which bind a
CC cerebral cortical voltage-dependent calcium channel alpha2delta subunit,
CC preferably alpha2delta-1 subunit. The method comprises contacting a
CC secreted soluble recombinant alpha2delta-1 subunit with a ligand of
CC interest and a labelled compound which binds the subunit, followed by
CC measuring the level of binding of the labelled compound to alpha2delta-1
CC subunit. The method is useful for screening ligands, preferably
CC biologically active products that modulate a nervous system function,
CC which bind a cerebral cortical voltage-dependent calcium channel
CC alpha2delta-1 subunit. The ligands identified by the method are useful
CC for treating disorders of the nervous system, including pain, epilepsy
CC and anxiety. The present sequence represents a human calcium channel
CC alpha2delta subunit related sequence.
XX
SQ Sequence 304 AA:

Query Match 47.7%; Score 41.5; DB 22: Length 304;
Best Local Similarity 69.2%; Pred. No. 1.6e+02;

DT		04-JUL-2001	(first entry)
XX			
DE	Human secreted soluble alpha2delta calcium channel subunit #9 protein.		
XX			
KW	Human; secreted calcium channel alpha2delta subunit; alpha2delta-2;		
KM	alpha2delta-3; alpha2delta-4; voltage-dependent calcium channel; VDCC;		
KW	gabapentin; scintillation proximity assay; SPA; nickel flashplate assay;		
KM	filter binding assay; wheat germ lectin flashplate assay.		
XX			
OS	Homo sapiens.		
XX			
PN	MO200119870-A2.		
XX			
PD	22-MAR-2001.		
XX			
PF	18-SEP-2000; 2000MO-EP09137.		
XX			
PR	16-SEP-1999; 99US-0397550.		
XX			
PA	(WARN) WARNER LAMBERT CO.		
XX			
F1	Brown JP, Bertelli F;		
XX			
XX	WPI; 2001-235262/24.		
DR	N-PSDB; AAS01412.		
XX			
PT	Calcium channel alpha2delta subunits, useful in e.g. SPA assays,		
PT	Flashplate assays, Nickel Flashplate assays, Filter binding assays or		
PT	wheat Germ Lectin Flashplate assays -		
XX			
PS	Disclosure: Page 86-87; 160pp; English.		
XX			
CC	The present sequence represents human secreted calcium channel		
CC	alpha2delta subunit #9 which is soluble and retains the functional		
CC	characteristics of the full length or wild type alpha2delta subunit		
CC	(AAU01025) from which it is derived. The invention relates to truncated		
CC	alpha2delta-2, alpha2delta-3 or alpha2delta-4 subunit soluble proteins		
CC	which retain their affinity for radioactively labeled gabapentin. The		
CC	alpha2delta subunit is 1 of the components of the heteromultimeric		
CC	voltage-dependent calcium channel (VDCC) complexes present in neuronal		
CC	and non-neuronal tissues including heart and skeletal muscle. Numerous		
CC	soluble forms of the human calcium channel alpha2delta subunits		
CC	(AAU01014-AAU01024 and AAU01032-AAU01038) and 5 soluble forms of the		
CC	porcine calcium channel alpha2delta subunits (AAU01027-AAU01031) are		
CC	described. The secreted soluble alpha2delta subunit may be used in assays		
CC	e.g. scintillation proximity assay (SPA), flashplate, nickel flashplate,		
CC	filter binding or wheat germ lectin flashplate assays to detect or		
CC	measure the binding or interaction of a ligand (e.g. gabapentin,		
CC	L-Norleucine, L-Allo-Isoleucine, L-methionine, L-leucine, L-isoleucine,		
CC	L-valine, Spermine and/or L-Phenylalanine) of a calcium channel		
CC	alpha2delta subunit.		
XX			
SQ	Sequence 350 AA:		
QY	4 I V P F L I L M G V T W C 16		
DG	I I : I I I I I I I I		
Db	209 I V I f l l e w s v - w g 220		
XX			
RESULT 45			
AAB62246			
ID	AAB62246 standard; Protein; 350 AA.		
AC	AAB62246; ✓		
XX			
DT	11-JUN-2001 (first entry)		
XX			
DE	Human calcium channel alpha2delta subunit related seq ID No. 18.		

XX Calcium channel alpha2delta subunit; alpha2delta-1; cerebral cortex;
 KW nervous system disorder; pain; epilepsy; anxiety; human.
 OS Homo sapiens.
 XX WO200120336-A2.
 XX
 XX 22-MAR-2001.
 PD
 PF 18-SEP-2000; 2000WO-EP09136.
 XX
 XX 16-SEP-1999; 99US-0397549.
 PR
 XX (WARN) WARNER LAMBERT CO.
 PA
 XX Bertelli F, Brown JP, Dissanayake V, Suman-Chauhan N, Gee NS;
 PI WPI: 2001-257902/26.
 XX DR N-PSDB; AAF57549.
 XX
 PT Competitive binding assay for screening ligands which bind a cerebral
 PT cortical voltage-dependent calcium channel alpha2-delta-1 subunit,
 PT where the ligands identified are useful for treating disorders of the
 PT nervous system, including pain -
 PS
 PS Disclosure; Page 86-87; 158pp; English.
 XX
 XX The invention relates to a new method for screening ligands which bind a
 CC cerebral cortical voltage-dependent calcium channel alpha2delta subunit,
 CC preferably alpha2delta-1 subunit. The method comprises contacting a
 CC secreted soluble recombinant alpha2delta-1 subunit with a ligand of
 CC interest and a labelled compound which binds the subunit, followed by
 CC measuring the level of binding of the labelled compound to alpha2delta-1
 CC subunit. The method is useful for screening ligands, preferably
 CC biologically active products that modulate a nervous system function,
 CC which bind a cerebral cortical voltage-dependent calcium channel
 CC alpha2delta-1 subunit. The ligands identified by the method are useful
 CC for treating disorders of the nervous system, including pain, epilepsy
 CC and anxiety. The present sequence represents a human calcium channel
 CC alpha2delta subunit related sequence.
 CC
 XX
 SO Sequence 350 AA:
 47.7%; Score 41.5; DB 22; Length 350;
 Best Local Similarity 69.2%; Pred. No. 1.8e+02;
 Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1.
 4 LVVFLLMGVYWG 16
 11:111 1 1 1
 209 LVLLFVSVWG 220

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2002, 16:49:55 ; Search time 19.31 Seconds

(without alignments)
18.646 Million cell updates/sec

Title: US-09-471-276-831_COPY_1_16

Perfect score: 87

Sequence: 1 MSMLVFLLMGVTWG 16

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Listing first 100 summaries

1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCtUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	50.6	280	4	US-09-264-419C-2
2	43	49.4	137	1	US-08-676-169-2
3	40	46.0	503	4	US-09-068-195-24
4	39	44.8	277	1	US-08-024-868-2
5	39	44.8	277	2	US-08-242-097-2
6	39	44.8	277	4	US-09-206-695-2
7	39	44.8	277	5	PCt-US96-11995-1
8	39	44.8	339	3	US-09-120-365-62
9	39	44.8	339	3	US-09-120-365-80
10	39	44.8	339	4	US-09-515-039-62
11	39	44.8	339	4	US-09-515-038-80
12	39	44.8	344	4	US-08-681-192-2
13	39	44.8	913	3	US-08-445-640-4
14	39	44.8	913	3	US-08-170-558-4
15	39	44.8	913	3	US-08-447-314-4
16	39	44.8	913	3	US-08-445-461-4
17	39	44.8	919	1	US-08-336-343A-2
18	39	44.8	1786	2	US-08-477-451-16
19	38.5	44.3	174	1	US-08-131-625B-14
20	38.5	44.3	174	2	US-08-799-464A-11
21	38.5	44.3	174	4	US-08-686-968C-10
22	38.5	44.3	174	4	US-09-113-750A-41
23	38.5	44.3	174	5	PCt-US95-09927-11
24	38.5	44.3	174	5	PCt-US95-10904-17
25	38.5	44.3	174	5	PCt-US95-10904-43
26	38.5	44.3	174	5	PCt-US95-10904-45
27	38.5	44.3	174	5	PCt-US95-10904-47

28	38.5	44.3	174	5	PCt-US95-10904-49	Sequence 49, Appl
29	38.5	44.3	174	5	PCt-US95-10904-51	Sequence 51, Appl
30	38.5	44.3	188	3	US-08-855-531D-37	Sequence 37, Appl
31	38.5	44.3	188	4	US-08-855-526B-37	Sequence 37, Appl
32	38	43.7	335	1	US-08-289-699A-4	Sequence 4, Appl
33	38	43.7	335	2	US-08-878-283-4	Sequence 4, Appl
34	38	43.7	384	3	US-08-852-824-4	Sequence 4, Appl
35	38	43.7	407	1	US-08-385-186-15	Sequence 15, Appl
36	38	43.7	414	1	US-08-614-801A-4	Sequence 4, Appl
37	38	43.7	425	1	US-08-385-186-13	Sequence 13, Appl
38	38	43.7	509	2	US-08-845-566-1	Sequence 1, Appl
39	37.5	43.1	271	2	US-08-790-374-2	Sequence 2, Appl
40	37.5	43.1	346	5	PCt-US96-10602-2	Sequence 2, Appl
41	37.5	43.1	2396	1	US-08-157-005-2	Sequence 2, Appl
42	37.5	43.1	2396	4	US-08-747-863-2	Sequence 2, Appl
43	37	42.5	286	5	PCt-US92-00282-9	Sequence 9, Appl
44	37	42.5	419	1	US-08-385-186-2	Sequence 2, Appl
45	37	42.5	494	3	US-09-006-636-8	Sequence 8, Appl
46	37	42.5	494	4	US-09-006-632-8	Sequence 8, Appl
47	37	42.5	531	5	PCt-US92-00282-5	Sequence 5, Appl
48	36	41.4	8	1	US-08-191-571-20	Sequence 20, Appl
49	36	41.4	8	5	PCt-US95-00296-20	Sequence 20, Appl
50	36	41.4	419	1	US-08-385-186-4	Sequence 4, Appl
51	36	41.4	421	2	US-08-484-993B-24	Sequence 24, Appl
52	36	41.4	421	2	US-08-484-158B-24	Sequence 24, Appl
53	36	41.4	421	2	US-08-484-596A-24	Sequence 24, Appl
54	36	41.4	421	2	US-08-480-150A-24	Sequence 24, Appl
55	36	41.4	421	2	US-08-458-711-24	Sequence 24, Appl
56	36	41.4	421	1	US-08-149-223A-24	Sequence 24, Appl
57	36	41.4	424	1	US-08-453-472-7	Sequence 7, Appl
58	36	41.4	424	1	US-08-038-948-4	Sequence 4, Appl
59	36	41.4	424	1	US-08-453-952-7	Sequence 7, Appl
60	36	41.4	424	2	US-08-862-903-7	Sequence 7, Appl
61	36	41.4	424	2	US-08-484-158B-61	Sequence 61, Appl
62	36	41.4	492	3	US-09-006-636-4	Sequence 4, Appl
63	36	41.4	492	4	US-09-006-632-4	Sequence 4, Appl
64	36	41.4	496	4	US-09-348-443-2	Sequence 2, Appl
65	36	41.4	518	4	US-09-147-009-4	Sequence 4, Appl
66	36	41.4	1724	2	US-08-477-441-15	Sequence 15, Appl
67	36	41.4	1968	1	US-08-455-533A-45	Sequence 45, Appl
68	36	41.4	1968	2	US-08-223-305C-45	Sequence 45, Appl
69	35.5	40.8	374	3	US-09-046-736-4	Sequence 4, Appl
70	35	40.2	12	3	US-09-186-260-10	Sequence 10, Appl
71	35	40.2	12	4	US-09-517-347-10	Sequence 10, Appl
72	35	40.2	129	4	US-09-107-383-13	Sequence 13, Appl
73	35	40.2	170	3	US-09-186-250-8	Sequence 8, Appl
74	35	40.2	170	4	US-09-517-347-4	Sequence 4, Appl
75	35	40.2	170	4	US-09-517-347-8	Sequence 8, Appl
76	35	40.2	178	3	US-09-081-180-2	Sequence 2, Appl
77	35	40.2	178	3	US-09-040-786-2	Sequence 2, Appl
78	35	40.2	178	3	US-08-467-822-44	Sequence 44, Appl
79	35	40.2	195	2	US-08-211-312-3	Sequence 3, Appl
80	35	40.2	195	2	US-08-472-285-3	Sequence 3, Appl
81	35	40.2	195	4	US-09-107-383-10	Sequence 10, Appl
82	35	40.2	195	4	US-08-433-697-44	Sequence 44, Appl
83	35	40.2	195	4	US-08-466-248-44	Sequence 44, Appl
84	35	40.2	195	4	US-08-472-929-3	Sequence 3, Appl
85	35	40.2	206	4	US-09-068-140A-8	Sequence 8, Appl
86	35	40.2	213	2	US-08-578-701A-1	Sequence 1, Appl
87	35	40.2	213	2	US-08-360-693-1	Sequence 1, Appl
88	35	40.2	213	4	US-09-107-383-14	Sequence 14, Appl
89	35	40.2	278	3	US-08-663-082-4	Sequence 4, Appl
90	35	40.2	303	2	US-08-846-762-20	Sequence 20, Appl
91	35	40.2	303	2	US-08-846-762-91	Sequence 91, Appl
92	35	40.2	318	2	US-08-562-114B-15	Sequence 15, Appl
93	35	40.2	318	4	US-09-109-205-17	Sequence 17, Appl
94	35	40.2	318	4	US-08-940-424-2	Sequence 2, Appl
95	35	40.2	318	4	US-08-729-594A-15	Sequence 15, Appl
96	35	40.2	318	4	US-08-729-594A-19	Sequence 39, Appl
97	35	40.2	365	2	US-08-979-424-3	Sequence 3, Appl
98	35	40.2	365	4	US-08-928-383B-2	Sequence 2, Appl
99	35	40.2	365	4	US-09-272-496-2	Sequence 2, Appl
100	35	40.2	365	4		

ALIGNMENTS

RESULT 1
US-09-264-419C-2
; Sequence 2, Application US/09264419C
; Patent No. 6174682
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: No. 6174682el Thiorodoxin Family Active Site Molecules and Uses
; FILE REFERENCE: MNI-076
; CURRENT APPLICATION NUMBER: US/09/264,419C
; CURRENT FILING DATE: 1999-03-08
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-264-419C-2

Query Match 50.6%; Score 44; DB 4; Length 280;
Best Local Similarity 53.8%; Pred. No. 28;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 MVLVFLLMGVTV 15
; : : : : :
DB 11 LAVVLLMGAPW 23

RESULT 2
US-08-676-169-2
; Sequence 2, Application US/08676169
; Patent No. 5773235
; GENERAL INFORMATION:
; APPLICANT: Chlrside, Ewan Douglas
; TITLE OF INVENTION: EQUINE ARTERITIS VIRUS PEPTIDES, ANTIBODIES
; TITLE OF INVENTION: AND THEIR USE IN A DIAGNOSTIC TEST
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHAYE P.C.
; STREET: 1100 NO. 5773235th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,169
; FILING DATE: 31-JUL-96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Crawford, Arthur R.
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 1498-85
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4006
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 137 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-676-169-2

Query Match 49.4%; Score 43; DB 1; Length 137;
Best Local Similarity 58.3%; Pred. No. 19;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSLVFLLMG 12
; : : : : :
DB 2 LSMIVLFLMG 13

RESULT 3
US-09-068-195-24
; Sequence 24, Application US/09068195B
; Patent No. 6140078
; GENERAL INFORMATION:
; APPLICANT: Sanders, Jan W.
; APPLICANT: Ledebner, Adrianus M.
; APPLICANT: Venema, Gerard
; APPLICANT: Kok, Jan
; TITLE OF INVENTION: Salt-Inducible Promoter Derivable from a Lactic Acid
; TITLE OF INVENTION: Bacterium, and Its Use in a Lactic Acid Bacterium for
; FILE REFERENCE: Sanders-60113/0252227
; CURRENT APPLICATION NUMBER: US/09/068,195B
; EARLIER FILING DATE: 1998-07-29
; EARLIER APPLICATION NUMBER: PCT/EP97/04755
; EARLIER FILING DATE: 1997-08-20
; EARLIER APPLICATION NUMBER: EP 97200744/7
; EARLIER FILING DATE: 1997-03-13
; EARLIER APPLICATION NUMBER: EP 96202444/4
; EARLIER FILING DATE: 1996-09-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 503
; TYPE: PRT
; ORGANISM: L. lactis MG1363
US-09-068-195-24

Query Match 46.0%; Score 40; DB 4; Length 503;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 LVVFLLMGVTV 15
; : : : : :
DB 35 LVVFLLMGGLW 46

RESULT 4
US-08-024-868-2
; Sequence 2, Application US/08024868
; Patent No. 5386013
; GENERAL INFORMATION:
; APPLICANT: Lee, Tae Ho
; APPLICANT: Wisniewski, Hans Georg
; APPLICANT: Vilcek, Jan
; TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Broadway and Nelmark
; STREET: 419 Seventh Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.24

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/024,868
3 FILING DATE:
4 CLASSIFICATION: 435
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: US/07/642,312
7 FILING DATE:
8 ATTORNEY/AGENT INFORMATION:
9 NAME: Livnat, Shmuel
10 REGISTRATION NUMBER: 33,949
11 REFERENCE/POCKET NUMBER: VILCEK-1
12 TELECOMMUNICATION INFORMATION:
13 TELEPHONE: 212-628-5197
14 TELEFAX: 212-737-3528
15 INFORMATION FOR SEQ ID NO: 2:
16 SEQUENCE CHARACTERISTICS:
17 LENGTH: 277 amino acids
18 TYPE: amino acid
19 TOPOLOGY: linear
20 MOLECULE TYPE: protein
21
22 US-08-024-868-2

```

Query Match	44.8%	Score 39;	DB 1;	length 277;
Best Local Similarity	50.0%;	Pred. No.	1.3e+02;	
Matches	8;	Conservative	3;	Mismatches 3;
				Indels 2;
				Gaps 1.

```
QY      3 MLVVFLLMGVT--WG 16
          :::| | | | | 1 11
Db      4 LIYFLLMEDTQGWG 19
```

```

: RESULT 5
: US-08-242-097-2
: Sequence 2, Application US/08242097
: Patent No. 5846763
: GENERAL INFORMATION:
: APPLICANT: Lee, Tae Ho
: APPLICANT: Wisniewski, Hans Georg
: APPLICANT: Vilcek, Jan
: TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding
: TITLE OF INVENTION: Therefor and Uses Thereof
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Broadway and Neilmark
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/242,097
: FILING DATE: 13-MAY-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: LEE26/VILCEK-1B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-628-5197
: TELEFAX: 212-737-3528
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 277 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-08-242-097-2

```

Query March	44.8%	Score 39;	DB 2;	Length 277;
Best Local Similarity	50.0%	Pred. No. 1.3e+02;		
Matches	8;	Conservative	3;	Mismatches 3;
				Indels 2;
				Gaps 1;
OY	3	MLVFLLLMGVT--WG	16	
	:	: : : : : : : :		
Db	4	LIVFLLLMEDTQGG	19	

```
QY      3 MLVFFLLMGVT--WG 16
          :::| | | | | | |
Db      4 LIYFLLEWEDTQGWG 19
```

```

RESULT      6
US-09-206-695-2
: Sequence 2, Application US/09206695
: Patent No. 6210905
: GENERAL INFORMATION:
: APPLICANT: Lee, Tae Ho
: APPLICANT: Wisniewski, Hans Georg
: APPLICANT: Vilcek, Jan
: TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding
: TITLE OF INVENTION: Therefore and Uses Thereof
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Broadway and Neimark
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/206,695
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/242,097
: FILING DATE: 13-MAY-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: LEE26/VILCEK-1B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-628-5197
: TELEFAX: 212-737-3528
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 277 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-09-206-695-2

Query Match      44.88; Score 39; DB 4; Length 277;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches      8; Conservative      3; Mismatches      3; Indels      2; Gaps      1;

OY      3 MIVFLIMGYT--WG 16
      :::||||| 1 11
DB      4 LIYFLLMEDYQGWG 19

RESULT      7
PCT-US96-11995-1
: Sequence 1, Application PC/TUS9611995
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: A Pharmaceutical Composition Containing TSG-6
: TITLE OF INVENTION: Protein for Treating Inflammatory Diseases and Cancer-Relat
: TITLE OF INVENTION: Methods of Using same
: NUMBER OF SEQUENCES: 9

```

```

OY      3  MUYVFLILMGVT--WG 16
        :::||||| 1 11
DB      4  LIYFLILMEDTQSG 19

RESULT  7
PCT-US96-11995-1
; Sequence 1, Application PC/FUS9611995
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: A Pharmaceutical Composition Containing TSG-6
; TITLE OF INVENTION: Protein for Treating Inflammatory Diseases and Cancer-Relat
; TITLE OF INVENTION: Methods of Using same
; NUMBER OF SEQUENCES: 9

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/11995
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/001,311
FILING DATE: 20-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: MISINMSKI-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-628-5197
TELEFAX: 212-737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 277 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-11995-1
```

```

Query Match      44.8%  Score 39;  DB 5;  Length 277;
Best Local Similarity 50.0%  Pred. No. 1.3e+02;
Matches      8;  Conservative      3;  Mismatches      3;  Indels      2;  Gaps      1;
```

```

Db      3  MCVFLLMGVT--WG 16
      4  LITFLLMEDTGGWG 19

RESULT      8
US-09-120-365-62
; Sequence 62, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natorl, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; EARLIER FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-120-365-62
```

```

Query Match      44.8%  Score 39;  DB 3;  Length 339;
Best Local Similarity 83.3%  Pred. No. 1.7e+02;
Matches      5;  Conservative      0;  Mismatches      1;  Indels      0;  Gaps      0;
```

```

QY      11  MCVTWG 16
      11  1111
Db      308  MCVTWG 313
```

```

RESULT      9
US-09-120-365-80
; Sequence 80, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natorl, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; EARLIER FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-120-365-80
```

```

Query Match      44.8%  Score 39;  DB 3;  Length 339;
Best Local Similarity 83.3%  Pred. No. 1.7e+02;
Matches      5;  Conservative      0;  Mismatches      1;  Indels      0;  Gaps      0;
```

```

QY      11  MCVTWG 16
      11  1111
Db      308  MCVTWG 313
```

```

RESULT      10
US-09-515-039-62
; Sequence 62, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natorl, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/515,039
; EARLIER FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Sarcophaga cathepsin L
US-09-515-039-62
```

```

Query Match      44.8%  Score 39;  DB 4;  Length 339;
Best Local Similarity 83.3%  Pred. No. 1.7e+02;
Matches      5;  Conservative      0;  Mismatches      1;  Indels      0;  Gaps      0;
```

```

QY      11  MCVTWG 16
      11  1111
Db      308  MCVTWG 313
```

```

RESULT      11
US-09-515-039-80
; Sequence 80, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natorl, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/515,039
; EARLIER FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
```


SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 80
LENGTH: 339
TYPE: PRT
ORGANISM: Sarcophaga cathepsin L
US-09-515-039-80

Query Match 44.8%; Score 39; DB 4; Length 339;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 11 MGVTWG 16
11 111
DB 308 MGTTWG 313

RESULT 12
US-08-681-192-2
Sequence 2, Application US/08681192
Patent No. 6287801
GENERAL INFORMATION:
APPLICANT: BERGSM, DEK
APPLICANT: ELSHOURBAGY, NABIL
APPLICANT: SARAU, HENRY
APPLICANT: RUBEN, STEVEN
TITLE OF INVENTION: G-PROTEIN COUPLED RECEPTOR HNRD578
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/681,192
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Schreck, Patricia A.
REGISTRATION NUMBER: 33,777
REFERENCE/DOCKET NUMBER: ATG50014
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5031
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
US-08-681-192-2

Query Match 44.8%; Score 39; DB 4; Length 344;
Best Local Similarity 60.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 SMUVFLLM 11
:::||||:|
DB 242 AIMVFLM 251

RESULT 13
US-08-445-640-4
Sequence 4, Application US/08445640
Patent No. 5709858
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Barton, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,640
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-445-640-4

Query Match 44.8%; Score 39; DB 1; Length 913;
Best Local Similarity 38.5%; Pred. No. 4.4e+02;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 3 MLVVFLLMGVTW 15
:::||||:|
DB 431 LILIALMLMRLHW 443

RESULT 14
US-08-170-558-4
Sequence 4, Application US/08170558
Patent No. 6001621
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-445-461-4

Query Match 44.8%; Score 39; DB 3; Length 913;
Best Local Similarity 38.5%; Pred. No. 4.4e+02;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 3 MLVFFLLMGVTW 15
:|:|:|:|
Db 431 LLIILMLRLHW 443

RESULT 17

US-08-336-343A-2
Sequence 2, Application US/08336343A
Patent No. 5677144
GENERAL INFORMATION:
APPLICANT: Ullrich, Axel
APPLICANT: Alves, Frauke
TITLE OF INVENTION: CCK-2, A No. 5677144el Receptor Tyrosine Kinase
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,343A
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7683-065
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 919 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-336-343A-2

Query Match 44.8%; Score 39; DB 1; Length 919;
Best Local Similarity 38.5%; Pred. No. 4.5e+02;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 3 MLVFFLLMGVTW 15
:|:|:|:|
Db 431 LLIILMLRLHW 443

RESULT 18
US-08-477-451-16
Sequence 16, Application US/08477451
Patent No. 5928865
GENERAL INFORMATION:

APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335,002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542

INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1786 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-451-16

Query Match 44.8%; Score 39; DB 2; Length 1786;
Best Local Similarity 40.0%; Pred. No. 8.7e+02;
Matches 6; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 MSMLVFFLLMGVTW 15
:|:|:|:|
Db 807 MRLLVFQMATIW 821

RESULT 19

US-08-131-625B-14
Sequence 14, Application US/08131625B
Patent No. 5695766
GENERAL INFORMATION:
APPLICANT: PAUL, PREM S.
APPLICANT: HALBUR, PATRICK G.
APPLICANT: MENG, XIANG-JIN
APPLICANT: LUM, MELISSA A.
APPLICANT: LYOO, YOUNG S.
TITLE OF INVENTION: VACCINES RAISING AN IMMUNOLOGICAL
TITLE OF INVENTION: RESPONSE AGAINST VIRUSES CAUSING PORCINE RESPIRATORY AND
TITLE OF INVENTION: REPRODUCTIVE DISEASES, METHODS OF PROTECTING A PIG AGAINST
TITLE OF INVENTION: A DISEASE CAUSED BY A RESPIRATORY AND REPRODUCTIVE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/131.625B
;; FILING DATE: 05-OCT-1993
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/969,071
;; FILING DATE: 30-OCT-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Lavalleye, Jean-Paul M.P.
;; REGISTRATION NUMBER: 31,451
;; REFERENCE/DOCKET NUMBER: 4625-016-55X CIP
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 413-3000
;; TELEFAX: (703) 413-2220
;; TELEX: 248855 OPAT UR
;; INFORMATION FOR SEQ ID NO: 14:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 174 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-131-625B-14

Query Match 44.3%; Score 38.5; DB 1; Length 174;
Best Local Similarity 45.0%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 MSMLVFFLLMGV-----TW 15
:::| | ||||| ||
Db 73 LTMGAVVALLMGVYSAIETW 92

RESULT 20
US-08-799-464A-11
; Sequence 11, Application US/08799464A
; Patent No. 5998601
; GENERAL INFORMATION:
; APPLICANT: Murlaugh, Michael P. et al.
; TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND
; TITLE OF INVENTION: METHODS OF USE
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John M. Collins
; STREET: 2405 Grand Blvd., Suite 400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: USA
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/799,464A
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/287,941
; FILING DATE: August 5, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26,262
; REFERENCE/DOCKET NUMBER: 22907
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816) 474-9050
; TELEFAX: (816) 474-9057
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 174 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

;; MOLECULE TYPE: protein
US-08-799-464A-11

Query Match 44.3%; Score 38.5; DB 2; Length 174;
Best Local Similarity 45.0%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 MSMLVFFLLMGV-----TW 15
:::| | ||||| ||
Db 73 LTMGAVVALLMGVYSAIETW 92

RESULT 21
US-08-686-968C-10
; Sequence 10, Application US/08686968C
; Patent No. 6221361
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; FILE REFERENCE: 39119-H/JML
; CURRENT APPLICATION NUMBER: US/08/686,968C
; CURRENT FILING DATE: 1996-07-25
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-08-686-968C-10

Query Match 44.3%; Score 38.5; DB 4; Length 174;
Best Local Similarity 45.0%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 MSMLVFFLLMGV-----TW 15
:::| | ||||| ||
Db 73 LTMGAVVALLMGVYSAIETW 92

RESULT 22
US-09-113-750A-41
; Sequence 41, Application US/09113750A
; Patent No. 6294176
; GENERAL INFORMATION:
; APPLICANT: David E. Junker and Mark D. Cochran
; TITLE OF INVENTION: Recombinant Raccoonpox Virus
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/113,750A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 55744
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)262-0400
; TELEFAX: (212)664-0525

TELEX: 422523
 INFORMATION FOR SEQ ID NO: 41:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 174 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-09-113-750A-41

Query Match 44.3%; Score 38.5; DB 4; Length 174;
 Best Local Similarity 45.0%; Pred. No. 1e+02;
 Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 MSMLVFLIMGV-----TW 15
 : : | | | | | | | |
 Db 73 LTMGAVVALLMGVSAIETW 92

RESULT 23
 PCT-US95-09927-11
 ; Sequence 11, Application PC/TUS9509927
 ; GENERAL INFORMATION:
 ; APPLICANT: Murtaugh, Michael P.
 ; TITLE OF INVENTION: VR-2332 VIRAL NUCLEOTIDE SEQUENCE AND
 ; NUMBER OF SEQUENCES: 26
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John M. Collins
 ; STREET: 1101 Walnut, Suite 1400
 ; CITY: Kansas City
 ; STATE: Missouri
 ; COUNTRY: USA
 ; ZIP: 64106
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/09927
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Collins, John M.
 ; REGISTRATION NUMBER: 26122
 ; REFERENCE/DOCKET NUMBER: 22907
 ; TELEPHONE: (816) 474-9050
 ; TELEFAX: (816) 474-9057
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 174 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US95-09927-11

Query Match 44.3%; Score 38.5; DB 5; Length 174;
 Best Local Similarity 45.0%; Pred. No. 1e+02;
 Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 MSMLVFLIMGV-----TW 15
 : : | | | | | | | |
 Db 73 LTMGAVVALLMGVSAIETW 92

RESULT 24
 PCT-US95-10904-17
 ; Sequence 17, Application PC/TUS9510904
 ; GENERAL INFORMATION:

APPLICANT: PAUL, PREM S.
 APPLICANT: MENG, XIANG-JIN
 APPLICANT: HALBUR, PATRICK G.
 APPLICANT: MOROZOV, IGOR
 APPLICANT: LUM, MELISSA A.
 TITLE OF INVENTION: A POLYNUCLEIC ACID ISOLATED FROM A
 TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV)
 TITLE OF INVENTION: A PROTEIN ENCODED BY THE POLYNUCLEIC ACID, A VACCINE
 TITLE OF INVENTION: PREPARED FROM OR CONTAINING THE POLYNUCLEIC ACID OR
 TITLE OF INVENTION: PROTEIN,
 NUMBER OF SEQUENCES: 77
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 ADDRESSEE: P.C.
 STREET: 1755 S. Jefferson Davis Highway, Suite 400
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/10904
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/131,625
 FILING DATE: 05-OCT-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Lavalleye, Jean-Paul M.P.
 REGISTRATION NUMBER: 31,451
 REFERENCE/DOCKET NUMBER: 4625-021-55X CIP
 TELEPHONE: (703) 413-3000
 TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 174 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-10904-17

Query Match 44.3%; Score 38.5; DB 5; Length 174;
 Best Local Similarity 45.0%; Pred. No. 1e+02;
 Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 MSMLVFLIMGV-----TW 15
 : : | | | | | | | |
 Db 73 LTMGAVVALLMGVSAIETW 92

RESULT 25
 PCT-US95-10904-43
 ; Sequence 43, Application PC/TUS9510904
 ; GENERAL INFORMATION:
 ; APPLICANT: PAUL, PREM S.
 ; APPLICANT: MENG, XIANG-JIN
 ; APPLICANT: HALBUR, PATRICK G.
 ; APPLICANT: MOROZOV, IGOR
 ; APPLICANT: LUM, MELISSA A.
 ; TITLE OF INVENTION: A POLYNUCLEIC ACID ISOLATED FROM A
 ; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV)
 ; TITLE OF INVENTION: A PROTEIN ENCODED BY THE POLYNUCLEIC ACID, A VACCINE
 ; TITLE OF INVENTION: PREPARED FROM OR CONTAINING THE POLYNUCLEIC ACID OR
 ; NUMBER OF SEQUENCES: 77
 ; CORRESPONDENCE ADDRESS:

```

ADDRESS: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10904
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/131,625
FILING DATE: 05-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-021-55X CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ. ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-10904-43

```

```

Query Match      44.3%; Score 38.5; DB 5; Length 174;
Best Local Similarity 45.0%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

```

```

QY 1 MSMLVFFLLMGV----TW 15
Db 73 LTMGAVALLMGVSAIETW 92

```

```

RESULT 26
PCT-US95-10904-45
Sequence 45, Application PC/TUS9510904
GENERAL INFORMATION:
APPLICANT: PAUL, PREM S.
APPLICANT: MENG, XIANG-JIN
APPLICANT: HALBUR, PATRICK G.
APPLICANT: MOROZOV, IGOR
APPLICANT: LUM, MELISSA A.
TITLE OF INVENTION: A POLYNUCLEIC ACID ISOLATED FROM A
TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV),
TITLE OF INVENTION: A PROTEIN ENCODED BY THE POLYNUCLEIC ACID, A VACCINE
TITLE OF INVENTION: PREPARED FROM OR CONTAINING THE POLYNUCLEIC ACID OR
TITLE OF INVENTION: PROTEIN,
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESSES:
ADDRESS: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10904
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/131,625
FILING DATE: 05-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-021-55X CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ. ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-10904-45

```

```

Query Match      44.3%; Score 38.5; DB 5; Length 174;
Best Local Similarity 45.0%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

```

```

QY 1 MSMLVFFLLMGV----TW 15
Db 73 LTMGAVALLMGVSAIETW 92

```

```

RESULT 27
PCT-US95-10904-47
Sequence 47, Application PC/TUS9510904
GENERAL INFORMATION:
APPLICANT: PAUL, PREM S.
APPLICANT: MENG, XIANG-JIN
APPLICANT: HALBUR, PATRICK G.
APPLICANT: MOROZOV, IGOR
APPLICANT: LUM, MELISSA A.
TITLE OF INVENTION: A POLYNUCLEIC ACID ISOLATED FROM A
TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV)
TITLE OF INVENTION: A PROTEIN ENCODED BY THE POLYNUCLEIC ACID, A VACCINE
TITLE OF INVENTION: PREPARED FROM OR CONTAINING THE POLYNUCLEIC ACID OR
TITLE OF INVENTION: PROTEIN,
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESSES:
ADDRESS: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10904
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/131,625
FILING DATE: 05-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-021-55X CIP
TELECOMMUNICATION INFORMATION:

```

TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-10904-47

Query Match 44.3%; Score 38.5; DB 5; Length 174;
Best Local Similarity 45.0%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

OY 1 MSMLVFLLMGV----TW 15
Db 73 LTMGAVVALLMGVSAIETW 92

RESULT 28
PCT-US95-10904-49

Sequence 49, Application PC/TUS9510904
GENERAL INFORMATION:

APPLICANT: PAUL, PREM S.

APPLICANT: MENG, XIANG-JIN

APPLICANT: HALBUR, PATRICK G.

APPLICANT: MOROZOV, IGOR

APPLICANT: LUM, MELISSA A.

TITLE OF INVENTION: A POLYNUCLEIC ACID ISOLATED FROM A

TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV),

TITLE OF INVENTION: A PROTEIN ENCODED BY THE POLYNUCLEIC ACID, A VACCINE

TITLE OF INVENTION: PREPARED FROM OR CONTAINING THE POLYNUCLEIC ACID OR

NUMBER OF SEQUENCES: 77

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10904

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/131,625

FILING DATE: 05-OCT-1993

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M.P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 4625-021-55X CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 49:

SEQUENCE CHARACTERISTICS:

LENGTH: 174 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US95-10904-49

Query Match 44.3%; Score 38.5; DB 5; Length 174;
Best Local Similarity 45.0%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

OY 1 MSMLVFLLMGV----TW 15
Db 73 LTMGAVVALLMGVSAIETW 92

RESULT 29
PCT-US95-10904-51

Sequence 51, Application PC/TUS9510904
GENERAL INFORMATION:

APPLICANT: PAUL, PREM S.

APPLICANT: MENG, XIANG-JIN

APPLICANT: HALBUR, PATRICK G.

APPLICANT: MOROZOV, IGOR

APPLICANT: LUM, MELISSA A.

TITLE OF INVENTION: A POLYNUCLEIC ACID ISOLATED FROM A

TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS (PRRSV)

TITLE OF INVENTION: A PROTEIN ENCODED BY THE POLYNUCLEIC ACID, A VACCINE

TITLE OF INVENTION: PREPARED FROM OR CONTAINING THE POLYNUCLEIC ACID OR

NUMBER OF SEQUENCES: 77

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10904

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/131,625

FILING DATE: 05-OCT-1993

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M.P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 4625-021-55X CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 51:

SEQUENCE CHARACTERISTICS:

LENGTH: 174 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US95-10904-51

Query Match 44.3%; Score 38.5; DB 5; Length 174;
Best Local Similarity 45.0%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

OY 1 MSMLVFLLMGV----TW 15
Db 73 LTMGAVVALLMGVSAIETW 92

RESULT 30

US-08-855-531D-37

Sequence 37, Application US/08855531D

```
Patent No. 6110467
GENERAL INFORMATION:
APPLICANT: PREM, PAUL S.
            HALBUR, PATRICK G.
            MENG, XIANG-JIN
            LIM, MELISSA A.
            LYOO, YOUNG S.
TITLE OF INVENTION: VACCINES RAISING AN IMMUNOLOGICAL
                    RESPONSE AGAINST VIRUSES CAUSING PORCINE RESPIRATORY AND
                    REPRODUCTIVE DISEASES, METHODS OF PROTECTING A PIG AGAINST
                    A DISEASE CAUSED BY A RESPIRATORY AND REPRODUCTIVE
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/855,531D
FILING DATE: 13-May-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/969,071
FILING DATE: 30-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: LAVALLEYE, JEAN-PAUL M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-038-55X DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-08-855-531D-37

Query Match          44.3% Score 38.5; DB 3; Length 188;
Best Local Similarity 45.0%; Pred No. 1.1e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 MSMLVFLMLGV----TW 15
:::| | ||||| ||
DB 87 LTMGAVVALLMGVSAIETW 106

RESULT 31
US-08-855-526B-37
Sequence 37, Application US/08855526B
Patent No. 6251404
GENERAL INFORMATION:
APPLICANT: PREM, PAUL S.
            HALBUR, PATRICK G.
            MENG, XIANG-JIN
            LIM, MELISSA A.
            LYOO, YOUNG S.
TITLE OF INVENTION: VACCINES RAISING AN IMMUNOLOGICAL
                    RESPONSE AGAINST VIRUSES CAUSING PORCINE RESPIRATORY AND
                    REPRODUCTIVE DISEASES, METHODS OF PROTECTING A PIG AGAINST
                    A DISEASE CAUSED BY A RESPIRATORY AND REPRODUCTIVE
```

```
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/855,526B
FILING DATE: 13-MAY-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/969,071
FILING DATE: 30-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: LAVALLEYE, JEAN-PAUL M.P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 4625-040-55X DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-855-526B-37

Query Match          44.3% Score 38.5; DB 4; Length 188;
Best Local Similarity 45.0%; Pred.No. 1.1e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

QY 1 MSMLVFLMLGV----TW 15
:::| | ||||| ||
DB 87 LTMGAVVALLMGVSAIETW 106

RESULT 32
US-08-289-699A-4
Sequence 4, Application US/08289699A
Patent No. 5655993
GENERAL INFORMATION:
APPLICANT: Fukudome, Kenji
            Esmon, Charles T.
TITLE OF INVENTION: Cloning and Regulation of an Endothelial
                    Cell Protein C/activated Protein C Receptor
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 West Peachtree
CITY: Street
STATE: Georgia
COUNTRY: US
ZIP: 30306-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,699A
```


FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRF152
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)873-8794
TELEFAX: (404)873-8795
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 335 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-289-699A-4

Query Match 43.7%; Score 38; DB 1; Length 335;
Best Local Similarity 50.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

OY 5 VVELLMGV--TWG 16
:::||||: ||
DB 4 LFLFLMALLQAMG 17

RESULT 33
US-08-878-283-4
Sequence 4, Application US/08878283
Patent No. 5852171
GENERAL INFORMATION:
APPLICANT: Fukudome, Kenji
APPLICANT: Esmon, Charles T.
TITLE OF INVENTION: Cloning and Regulation of an Endothelial
TITLE OF INVENTION: Cell Protein C/activated Protein C Receptor
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 West Peachtree
STREET: Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30306-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,283
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,699
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRF152
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)873-8794
TELEFAX: (404)873-8795
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 335 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-878-283-4

Query Match 43.7%; Score 38; DB 2; Length 335;
Best Local Similarity 50.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

OY 5 VVELLMGV--TWG 16
:::||||: ||
DB 4 LFLFLMALLQAMG 17

RESULT 34
US-08-852-824-4
Sequence 4, Application US/08852824C
Patent No. 6060272
GENERAL INFORMATION:
APPLICANT: Li et al.
TITLE OF INVENTION: Human G-Protein Coupled Receptors
FILE REFERENCE: 1488,1220000
CURRENT APPLICATION NUMBER: US/08/852,824C
CURRENT FILING DATE: 1997-05-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 384
TYPE: PRT
ORGANISM: genomic
US-08-852-824-4

Query Match 43.7%; Score 38; DB 3; Length 384;
Best Local Similarity 37.5%; Pred. No. 2.6e+02;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 MSLVFLMGVTVWG 16
:::||||: |
DB 254 LMLLAFLVCMGPLFG 269

RESULT 35
US-08-385-186-15
Sequence 15, Application US/08385186
Patent No. 5744594
GENERAL INFORMATION:
APPLICANT: Agelman, John P
APPLICANT: Bond, Chris T
TITLE OF INVENTION: DNA ENCODING ATP-SENSITIVE POTASSIUM
TITLE OF INVENTION: CHANNEL PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourie and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/385,186
FILING DATE: 07-FEB-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/288,510
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/193,372
FILING DATE: 08-FEB-1994
ATTORNEY/AGENT INFORMATION:

NAME: Kruse, No. 5744594man J.
REGISTRATION NUMBER: 35,235
REFERENCE/DOCKET NUMBER: 14210-2-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 407 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-385-186-15

Query Match 43.7% Score 38; DB 1; Length 407;
Best Local Similarity 41.7% Pred. No. 2.7e+02;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 4 LVFLLMGVTW 15
|::|::|::|
Db 77 LLIFFVWVTVTW 88

RESULT 36
US-08-614-801A-4
Sequence 4, Application US/08614801A
Patent No. 57443324
GENERAL INFORMATION:
APPLICANT: Lester, Henry A.
APPLICANT: Davidson, No. 5744324man
APPLICANT: Kofuji, Paulo
TITLE OF INVENTION: INWARD RECTIFIER, G-PROTEIN ACTIVATED,
TITLE OF INVENTION: MAMMALIAN, POTASSIUM CHANNELS AND USES THEREOF
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614.801A
FILING DATE: 07-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/066,371
FILING DATE: 21-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-63098/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-614-801A-4

Query Match 43.7% Score 38; DB 1; Length 414;
Best Local Similarity 41.7% Pred. No. 2.8e+02;

Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
QY 4 LVFLLMGVTW 15
|::|::|::|
Db 95 LLIFFVWVTVTW 106

RESULT 37
US-08-385-186-13
Sequence 13, Application US/08385186
Patent No. 5744594
GENERAL INFORMATION:
APPLICANT: Adelman, John P.
APPLICANT: Ashford, Michael J.
APPLICANT: Bond, Chris P.
TITLE OF INVENTION: DNA ENCODING ATP-SENSITIVE POTASSIUM
TITLE OF INVENTION: CHANNEL PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Knourie and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/385,186
FILING DATE: 07-FEB-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/288,510
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/193,372
FILING DATE: 08-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kruse, No. 5744594man J.
REGISTRATION NUMBER: 35,235
REFERENCE/DOCKET NUMBER: 14210-2-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 425 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-385-186-13

Query Match 43.7% Score 38; DB 1; Length 425;
Best Local Similarity 41.7% Pred. No. 2.9e+02;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 4 LVFLLMGVTW 15
|::|::|::|
Db 95 LLIFFVWVTVTW 106

RESULT 38
US-08-845-566-1
Sequence 1, Application US/08845566
Patent No. 5912144
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Guegler, Karl
TITLE OF INVENTION: EDC-1 LIKE RECEPTOR

NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/845,566
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PR-0271 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 509 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TYMNOR01
CLONE: 144690
US-08-845-566-1

Query Match 43.7%; Score 38; DB 2; Length 509;
Best Local Similarity 37.5%; Pred. No. 3.4e+02;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSMLVFFLLMGVTWG 16
DB 254 LMLLAFVCMGPLFG 269

RESULT 39
US-08-790-374-2
Sequence 2, Application US/08790374
GENERAL INFORMATION:
PATENT NO. 5863734
APPLICANT: Karayiorjou, Maria
APPLICANT: Gogos, Joseph A.
TITLE OF INVENTION: METHODS OF TREATMENT FOR OBSESSIVE-COMPULSIVE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauders & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/790,374
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-189
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 271 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-790-374-2

Query Match 43.1%; Score 37.5; DB 2; Length 271;
Best Local Similarity 58.8%; Pred. No. 2.1e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 5; Gaps 2;

QY 3 MLVFFLL--WGVTWG 16
DB 18 LLVVLRLRHMG--WG 32

RESULT 40
PCT-US96-10602-2
Sequence 2, Application PC/TUS9610602
GENERAL INFORMATION:
APPLICANT: The General Hospital Corporation
TITLE OF INVENTION: INHIBITION OF HEPATITIS B REPLICATION
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10602
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017,814
FILING DATE: 20-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/282001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 346 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-10602-2

Query Match 43.1%; Score 37.5; DB 5; Length 346;
Best Local Similarity 41.2%; Pred. No. 2.7e+02;

Matches 7; Conservative 5; Mismatches 4; Indels 1; Gaps 1;
Qy 1 MSLVFFLLMVCV-TWC 16
: | : | | : | |
Db 306 ISLAFLLMIMFWG 322

RESULT 41
US-08-157-005-2
; Sequence 2, Application US/08157005
; Patent No. 5620691
; GENERAL INFORMATION:
; APPLICANT: Mensvoort, Gert
; APPLICANT: Terpstra, Catharinus
; APPLICANT: Pol, Johannes M
; APPLICANT: Moorman, Robertus J
; APPLICANT: Meulenbergh, Johanna J
; TITLE OF INVENTION: CAUSITIVE AGENT OF THE MYSTERY SWINE DISEASE,
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/157,005
; FILING DATE: 26-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91201398.4
; FILING DATE: 06-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92200781.0
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL92/00096
; FILING DATE: 05-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Moran, Thomas F
; REGISTRATION NUMBER: 16,579
; REFERENCE/DOCKET NUMBER: 44819
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 315-1931
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2396 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-157-005-2

Query Match 43.1%; Score 37.5; DB 1; Length 2396;
Best Local Similarity 44.4%; Pred. No. 1.9e+03;
Matches 8; Conservative 2; Mismatches 5; Indels 3; Gaps 1;

Qy 1 MSLVFFLLMVCV-TWC 15
: | : | | : | |
Db 1902 VOLLCVFFLLMIMFWG 1919

RESULT 42
US-08-747-863-2

; Sequence 2, Application US/08747863
; Patent No. 6197310
; GENERAL INFORMATION:
; APPLICANT: Mensvoort, Gert
; APPLICANT: Terpstra, Catharinus
; APPLICANT: Pol, Johannes M
; APPLICANT: Moorman, Robertus J
; APPLICANT: Meulenbergh, Johanna J
; TITLE OF INVENTION: CAUSITIVE AGENT OF THE MYSTERY SWINE DISEASE,
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Trask, Britt & Rossa
; STREET: 525 South 300 East
; CITY: Salt Lake City
; STATE: Utah
; COUNTRY: USA
; ZIP: 84111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS95
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,863
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/157,005
; FILING DATE: 26-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91201398.4
; FILING DATE: 06-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92200781.0
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL92/00096
; FILING DATE: 05-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Turner, Allen C.
; REGISTRATION NUMBER: 33041
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (801) 532-1922
; TELEFAX: (801) 531-9168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2396 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-747-863-2

Query Match 43.1%; Score 37.5; DB 4; Length 2396;
Best Local Similarity 44.4%; Pred. No. 1.9e+03;
Matches 8; Conservative 2; Mismatches 5; Indels 3; Gaps 1;

Qy 1 MSLVFFLLMVCV-TWC 15
: | : | | : | |
Db 1902 VOLLCVFFLLMIMFWG 1919

RESULT 43
PCT-US92-00282-9
; Sequence 9, Application PC/TUS9200282
; GENERAL INFORMATION:
; APPLICANT: OWENS, IDA S.
; APPLICANT: RITTER, JOSEPH K.
; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
; THEREIN.
; NUMBER OF SEQUENCES: 40

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
;; STREET: 1615 L STREET, N.W.
;; CITY: WASHINGTON
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20036-5601
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US92/00282
;; FILING DATE: 19920110
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: SCOTT, WATSON T.
;; REGISTRATION NUMBER: 26581
;; REFERENCE/DOCKET NUMBER: 91532-PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-861-3000
;; TELEFAX: 202-822-0944
;; TELEX: 6714627 CUSH
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 286 amino acids
;; TYPE: AMINO ACID
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; PCT-US92-00282-9

Query Match 42.5%; Score 37; DB 5; Length 286;
Best Local Similarity 58.3%; Pred. No. 2.6e+02;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 5 VVFLLMGVTVG 16
| | | | |
DB 15 VVFLALMGVVG 26
RESULT 44
US-08-385-186-2
; Sequence 2, Application US/08385186
; Patent No. 5744594
; GENERAL INFORMATION:
; APPLICANT: Adelman, John P
; APPLICANT: Ashford, Michael J
; TITLE OF INVENTION: DNA ENCODING ATP-SENSITIVE POTASSIUM
; TITLE OF INVENTION: CHANNEL PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Knoutle and Crew
; STREET: Stewart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/385,186
; FILING DATE: 07-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/288,510
; FILING DATE: 10-AUG-1994

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/193,372
;; FILING DATE: 08-FEB-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kruse, No. 5744594man J.
;; REGISTRATION NUMBER: 35,235
;; REFERENCE/DOCKET NUMBER: 14210-2-2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 543-9600
;; TELEFAX: (415) 543-5043
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 419 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-385-186-2

Query Match 42.5%; Score 37; DB 1; Length 419;
Best Local Similarity 50.0%; Pred. No. 3.9e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 4 LVVFLMGVTV 15
| | | | |
DB 90 LVVFTMVTYVTV 101

RESULT 45
US-09-006-636-8
; Sequence 8, Application US/09006636
; Patent No. 6005092
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shani, Ziv
; TITLE OF INVENTION: ARABIDOPSIS THALIANA ENDO-1,4-BETA-
; TITLE OF INVENTION: GLUCANASE GENE, PROMOTER AND PROTEIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/006,636
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7809-019
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;; US-09-006-636-8
Query Match 42.5%; Score 37; DB 3; Length 494;

Best Local Similarity 71.4%; Pred. No. 4.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 9 LLMGYTW 15
1111 : 1
Db 241 LLMGASW 247

Search completed: January 7, 2002, 16:49:57
Job time: 284 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2002, 16:52:06 ; Search time 19.35 Seconds
(without alignments)
62.987 Million cell updates/sec

Title: US-09-471-276-831_COPY_1_16
Perfect score: 87
Sequence: 1 MSMLVFLLMGVTWG 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR_68:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	57.5	98	2	H75494 hypothetical prote
2	50	57.5	243	2	E72405 ABC transporter, p
3	49	56.3	217	2	S01095 hypothetical prote
4	49	56.3	1107	2	B86112 probable periplasm
5	49	56.3	1107	2	E65226 hypothetical 123.8
6	48	55.2	471	2	E83173 hypothetical prote
7	47	54.0	48	2	D69402 hypothetical prote
8	47	54.0	300	2	D83158 hypothetical prote
9	47	54.0	389	2	H82825 transport protein
10	46	52.9	374	2	F70100 transpore protein
11	45.5	52.3	448	2	A83498 two-component sens
12	45	51.7	275	2	E75548 conserved hypotnet
13	45	51.7	436	2	T36320 hypothetical prote
14	45	51.7	522	2	T21591 hypothetical prote
15	44.5	51.1	227	2	A82542 conserved hypotnet
16	44	50.6	126	2	S77163 hypothetical prote
17	44	50.6	287	2	T12471 hypothetical prote
18	44	50.6	298	2	F81301 probable integral
19	44	50.6	323	2	T35734 ADP,ATP aminopept
20	44	50.6	379	2	T04608 ADP,ATP carrier pr
21	44	50.6	417	2	H83708 hypothetical prote
22	44	50.6	486	2	T49017 hypothetical prote
23	44	50.6	688	2	H96681 protein P1E2.10 f
24	44	50.6	1197	2	D82696 conserved hypotnet
25	43.5	50.0	305	2	F70030 conserved hypotnet
26	43	49.4	103	2	S77270 hypothetical prote
27	43	49.4	255	2	F39925 hypothetical prote
28	43	49.4	305	1	S52775 hypothetical prote
29	43	49.4	308	1	S30259 ADP,ATP carrier pr

30	43	49.4	402	2	S51791 Rhodopsin (similar
31	43	49.4	448	1	S29483 Rhodopsin (similar
32	43	49.4	621	2	H84922 Rhodopsin (similar
33	43	49.4	710	2	T05361 Rhodopsin (similar
34	43	49.4	3345	2	T13423 Rhodopsin (similar
35	42.5	48.9	218	2	F81111 Rhodopsin (similar
36	42.5	48.9	218	2	H81906 Rhodopsin (similar
37	42	48.3	102	2	H25035 Rhodopsin (similar
38	42	48.3	102	2	G25035 Rhodopsin (similar
39	42	48.3	338	2	E72636 Rhodopsin (similar
40	42	48.3	406	2	A83795 Rhodopsin (similar
41	42	48.3	407	2	T36979 Rhodopsin (similar
42	42	48.3	438	1	S24065 Rhodopsin (similar
43	42	48.3	455	1	OOOCC Rhodopsin (similar
44	42	48.3	499	2	S28306 Rhodopsin (similar
45	42	48.3	511	2	G64902 Rhodopsin (similar
46	42	48.3	511	2	H85726 Rhodopsin (similar
47	42	48.3	543	2	A28671 Rhodopsin (similar
48	42	48.3	615	2	T43330 Rhodopsin (similar
49	41.5	47.7	113	2	T45276 Rhodopsin (similar
50	41.5	47.7	793	2	C83260 Rhodopsin (similar
51	41	47.1	129	2	JC5902 Rhodopsin (similar
52	41	47.1	181	2	A28782 Rhodopsin (similar
53	41	47.1	241	2	C45271 Rhodopsin (similar
54	41	47.1	257	2	F86697 Rhodopsin (similar
55	41	47.1	312	2	D86085 Rhodopsin (similar
56	41	47.1	312	2	H78664 Rhodopsin (similar
57	41	47.1	313	2	H71341 Rhodopsin (similar
58	41	47.1	348	2	T21627 Rhodopsin (similar
59	41	47.1	384	2	D75201 Rhodopsin (similar
60	41	47.1	436	2	T36480 Rhodopsin (similar
61	41	47.1	525	2	T01108 Rhodopsin (similar
62	41	47.1	546	1	B24707 Rhodopsin (similar
63	41	47.1	576	2	T11046 Rhodopsin (similar
64	41	47.1	1165	2	S58236 Rhodopsin (similar
65	41	47.1	1234	2	T17515 Rhodopsin (similar
66	40	46.0	67	2	T17761 Rhodopsin (similar
67	40	46.0	167	2	F83845 Rhodopsin (similar
68	40	46.0	190	2	S20973 Rhodopsin (similar
69	40	46.0	190	2	F85901 Rhodopsin (similar
70	40	46.0	202	2	T29136 Rhodopsin (similar
71	40	46.0	218	2	H75155 Rhodopsin (similar
72	40	46.0	259	2	S60882 Rhodopsin (similar
73	40	46.0	305	2	S68154 Rhodopsin (similar
74	40	46.0	321	2	T27521 Rhodopsin (similar
75	40	46.0	337	2	E84133 Rhodopsin (similar
76	40	46.0	354	1	A48191 Rhodopsin (similar
77	40	46.0	364	2	C86015 Rhodopsin (similar
78	40	46.0	365	2	A85930 Rhodopsin (similar
79	40	46.0	376	2	S47693 Rhodopsin (similar
80	40	46.0	431	2	D83160 Rhodopsin (similar
81	40	46.0	436	2	A69777 Rhodopsin (similar
82	40	46.0	446	2	H65060 Rhodopsin (similar
83	40	46.0	451	1	S27617 Rhodopsin (similar
84	40	46.0	455	1	A69753 Rhodopsin (similar
85	40	46.0	462	2	G75100 Rhodopsin (similar
86	40	46.0	503	2	C86786 Rhodopsin (similar
87	40	46.0	660	2	A64739 Rhodopsin (similar
88	40	46.0	660	2	B85499 Rhodopsin (similar
89	39.5	45.4	307	2	C84962 Rhodopsin (similar
90	39.5	45.4	455	2	P86573 Rhodopsin (similar
91	39.5	45.4	455	2	A72050 Rhodopsin (similar
92	39.5	45.4	456	2	A71501 Rhodopsin (similar
93	39	44.8	118	2	S76608 Rhodopsin (similar
94	39	44.8	157	2	C70080 Rhodopsin (similar
95	39	44.8	178	2	A69841 Rhodopsin (similar
96	39	44.8	183	2	T59442 Rhodopsin (similar
97	39	44.8	218	2	S67481 Rhodopsin (similar
98	39	44.8	220	2	I80329 Rhodopsin (similar
99	39	44.8	236	2	C83888 Rhodopsin (similar
100	39	44.8	239	2	T36765 Rhodopsin (similar

A:Reference number: A64720; MUID:97426617
A:Accession: E65226
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1107 <BLAT>
A:Cross-references: GB:AE000488; GB:U00096; NID:g2367354; PIDN:ANC77119.1; PID:g2367355
A:Experimental source: GB:AE000488; strain K-12; substrain M6165
R:Butland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A:Title: Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92
A:Reference number: S56314; MUID:95334362
A:Accession: S56387
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-746, 'X', 748-1014, 'A', 1016-1107 <BUR>
A:Cross-references: EMBL:U014003; NID:q1263172; PIDN:AA97058.1; PID:g537003
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
C:Genetics:
A:Gene: yJep
A:Start codon: GTG

Query Match	56.3%	Score 49	DB 2	Length 1107
Best Local Similarity	37.5%	Pred. No. 20		
Matches	6	Conservative	6	Mismatches 4
				Indels 0
				Gaps 0
QY	1	MSMLVFFLLGCVWG	16	
	:::	:	::	
DB	1	MRLLITFLMWCISWG	16	

RESULT 6
E83173

hypothetical protein PA3789 [imported] - Pseudomonas aeruginosa (strain PA01)
C.Species: Pseudomonas aeruginosa
C.Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C.Accession: E83173
R.Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim
: Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A.Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A.Reference number: A82950; MUID:20437337
A.Accession: E83173
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-471 <STO>
A.Cross-references: GB:AE004797; GB:AE004091; NID:g9949950; PIDN:AA607176.1; GSPDB:GM009
A.Experimental source: strain PA01
C.Genetics:
A.Gene: PA3789
C:Superfamily: Neisseria meningitidis hypothetical protein NMB1721

Query Match	55.28	Score 48	DB 2	length 471
Best Local Similarity	57.18	Pred. No. 14		
Matches	8	Conservative	5	Mismatches 1
				Indels 0
				Gaps 0
QY	2	SMLVVFLILMGVTM	15	.
		1:1:1:1:1:1:1:1		
Db	201	SLLLFLMLSGMTM	214	

RESULT 7
D69402
hypothetical protein AF1221 - Archaeoglobus fulgidus
C.Species: Archaeoglobus fulgidus
C.Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C.Accession: D69402
R.Klek, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirnness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch.
A:Reference number: A69250; MUID:96049343
A:Accession: D69402
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-48 <KLE>
A:Cross-references: GB:AE001020; GB:AE000782; NID:92689343; PIDN:AAE90036.1; PID:g2644

Query Match	54.0%;	Score 47;	DB 2;	Length 48;
Best Local Similarity	54.5%;	Pred. No. 3.4;		
Matches	6;	Conservative	4;	Mismatches 1;
				Indels 0;
				Gaps 0;

```
QY      5 VFLLMGVFW 15
          :| | | :| | :|
Db     37 IVLLIWGLSW 47
```

hypothetical protein PA3897 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: DB3158
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.. Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: DB3158
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-300 <STO>
A:Cross-references: GB:AE004807; GB:AE004091; NID:g9950076; PIDN:AA607284.1; GSPDB:GN0
A:Experimental source: Strain PA01
C:Genetics:
A:Gene: PA3897

Query	Match	Similarity	Score	DB	Length	300;
Best	Local	Similarity	40.0%;			
Matches	6;	Conservative	6;	Mismatches	3;	Indels
					0;	Gaps
QY	1	MSMLVVELLWCVTW	15			
	:	1	:	:	:	:
	:	1	:	:	:	:
Db	3	LSLYLLTLVILWGTW	17			

RESULT	9
H82825	
transport protein XF0281 [imported] - Xylella fastidiosa (strain 9a5c)	

C.Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C.Accession: H82825
R.anonymus, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A.Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A.Reference number: A82515; MUID:20365717
A.Note: for a complete list of authors see reference number A59328 below

A:Residues: 1389 <SIM>
A:Cross-references: GB:AE003881; GB:AE003849; NID:9105093; PIDN:AAFP83094.1; GSPDB:GN8
A:Experimental source: straln 9A5C
R:Simston, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvaranga, R.
B:Jones, M.R.S.; Bueno, M.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreiro
as Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.: Junqueira, M.L.; Kemper, E.L.; Kitahara, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigle-Chado, M.A.; Medeira, A.M.B.N.; Medeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B.A.; Augusto: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaka, A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tsunako, M.H.; Valada, H.; Van Sluys, M.A.; Verjovsky-Almeida, S.; Vettore, A.L.; A:Reference: number: A59328
C:Contents: annotation
Gene: XF0281

Query Match	54.0%	Score 47	DB 2	Length 389
Best Local Similarity	64.3%	Score No. 17		
Matches 9	Conservative	2	Mismatches 3	Indels 0
Gaps				0
QY	3	MLVFFLLMGVTWG	16	
	1111111111			
DB	344	MLIVALLLFGSLMG	357	

RESULT 10
E70100
conserved hypothetical integral membrane protein BB0006 - Lyme disease spirochete
C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 17-Mar-2000
C:Accession: F70100
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whitton, D.; Peterson, J.; Karlavag, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vogt, B.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.
A:Reference number: A70100; MUID:98065943
A:Accession: F70100
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-574 <KLEP>
A:Cross-references: GB:AE001115, GB:AE000783, NID:g2687879, PIDN:AC66397.1; PID:g2687880
A:Experimental source: strain B31
A:Superfamily: conserved hypothetical protein HT0338

Query Match	52.98	Score 46	DB 2	Length 374
Best Local Similarity	50.08	Pred No	23	
Matches 7	Conservative 2	Mismatches 5	Indels 0	Gaps 0
QY	3 MLVFLIMGVTWG	16		
	:			
DB	311 LILCLPLFPGWLMG	324		

```

RESULT 11
AB34348
Two-component sensor Phoc PA1180 [Imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: AB34348
R:Stover, C.K.; Pham, X.O.; Ewlin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Ras, A.; Lardig, K.; Lim,
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437337
A:Accession: AB34348
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-448 <STC>
A:Cross-references: GB:AE004548; GB:AE004091; NID:99947102; PIDN:AM04569.1; GSPDB:GM001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: phoc; PA1180

```

	Query Match	Score	DB	length
Best Local Similarity	56.2%	Pred. No.	30:	
Matches	9:	Conservative	4:	Mismatches
OY	2	SMALVFFLLIM-CVTWG	16	
	:::::		::	
Db	173	ALLVALGLLGLTWG	188	

RESULT 12
E75548
conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)
C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: E75548
R:White, O.; Eisten, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; MUID:20036896
A:Accession: E75548
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-275 <WHI>
A:Cross-references: GH:AE001882; GH:AE000513; NID:96457865; PIDN:AA09788.1; PID:964545
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0203
A:Map position: 1

Query Match	51.7%	Score 45	DB 2	Length 275
Best Local Similarity	66.7%	Pred. No. 24		
Matches 8	Conservative 2	Mismatches 2	Indels 0	Gaps 0
QY	1	MSMLVVELLWG	12	
	:			
DB	135	LPMLVLFALLWG	146	

RESULT 13
T36320
Hypothetical protein SCE9.01 - Streptomyces coelicolor (fragment)
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36320
R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999
A:Reference number: 221575
A:Accession: T36320
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-436 <OL1>
A:Cross-references: EMBL:AL049841; PIDN:CAB42747.1; GSPDB:GN00070; SCOEDB:SCE9.01
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE9.01

```

Query Match          51.7%;   Score 45;   DB 2;   Length 436;
Best Local Similarity 64.3%;   Pred. No. 35;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 MLVFFLLMGVTWG 16
   : 1 1 1 1 1 1 1
Db 179 VLAVSLLLTGATWG 192

RESULT 14
T21591

```

hypothetical protein F31C3.6 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000

C:Accession: J21591

R:Collage, A.

submitted to the EMBL Data Library, March 1997

A:Reference number: 219446

A:Accession: J21591

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-522 <WIL>

A:Cross-references: EMBL:293784; PIDN:CA807196.2; GSPDB:GN00019; CESP:F31C3.6

A:Experimental source: clone F31C3

C:Genetics:

A:Gene: CESP:F31C3.6

A:Map position: 1

A:Introns: 41/1; 81/3; 105/1; 130/3; 204/1; 274/3; 309/1; 430/1; 488/2

C:Superfamily: *Caenorhabditis elegans* hypothetical protein F31C3.6

Query Match 51.7%; Score 45; DB 2; Length 522;

Best Local Similarity 35.7%; Pred. No. 40;

Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 3 MLVFLILMGVTWG 16

Db 238 VLITLVVWNICWG 251

RESULT 15

A82542 conserved hypothetical protein XF2574 [imported] - *Xylella fastidiosa* (strain 9a5c)

C:Species: *Xylella fastidiosa*

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: A82542

R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: A82542

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-227 <STM>

A:Cross-references: GB:AE004064; GB:AE003849; NID:g9107775; PIDN:AAF5371.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincanl, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fromm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, F

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracosta, E.C.; Miyaki, C.Y.

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2574

Query Match 51.1%; Score 44.5; DB 2; Length 227;

Best Local Similarity 62.5%; Pred. No. 25;

Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

OY 1 MSMLVFLILMGVTWG 15

Db 153 MAMLTVLVLLMDVRW 168

RESULT 16

S77183

hypothetical protein sl11698 - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: S77183

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,

O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*

s.

A:Reference number: S74322; MUID:97061201

A:Accession: S77183

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-126 <KAN>

A:Cross-references: EMBL:D09098; GB:AB001339; NID:q1652725; PIDN:BA017741.1; PID:9165

C:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: *Synechocystis* hypothetical protein sl11698

Query Match 50.6%; Score 44; DB 2; Length 126;

Best Local Similarity 57.1%; Pred. No. 19;

Matches 12; Conservative 1; Mismatches 2; Indels 6; Gaps 2;

OY 1 MSMLVFLILMGVTWG 16

Db 78 MSIL-PELILGALCMNVKMG 97

RESULT 17

T12471

hypothetical protein DKFZP564E1962.1 - human (fragment)

C:Species: *Homo sapiens* (man)

C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 23-Jul-1999

C:Accession: T12471

R:Blum, H.; Baerends, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, June 1999

A:Reference number: T12471

A:Accession: T12471

A:Status: preliminary

A:Molecule type: BLND

A:Residues: 1-287 <BLND>

A:Cross-references: EMBL:AL080080

A:Experimental source: fetal brain; clone DKFZP564E1962

C:Genetics:

A:Note: DKFZP564E1962.1

Query Match 50.6%; Score 44; DB 2; Length 287;

Best Local Similarity 53.8%; Pred. No. 35;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 MLVFLILMGVTWG 15

Db 18 LAVLVLLMGAPW 30

RESULT 18

F81301

probable integral membrane protein Cj1544c [imported] - *Campylobacter jejuni* (strain

C:Species: *Campylobacter jejuni*

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000

C:Accession: F81301

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chl

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Ba

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals

A:Reference number: A81250; MUID:20150912

A:Accession: F81301

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-298 <PAR>
A:Cross-references: GB:AL139078; GB:AL111168; NID:q6968723; PIDN:CAB73960.1; PID:q696896
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj1544c

Query Match 50.6%; Score 44; DB 2; Length 298;
Best Local Similarity 46.2%; Pred. No. 36;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 3 MLVYFLLMGVTW 15
:|:|:|:|:|:|
Db 10 LVIAMFLMCSSW 22

RESULT 19
T35734
Probable aminopeptidase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
C:Accession: T35734
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, August 1999
A:Reference number: 221588
A:Accession: T35734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-323 <SAND>
A:Cross-references: EMBL:AL109732; PIDN:CA852045.1; GSPDB:GN00070; SCOEDB:SC7H2.03c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC7H2.03c
C:Superfamily: proline aminopeptidase

Query Match 50.6%; Score 44; DB 2; Length 323;
Best Local Similarity 75.0%; Pred. No. 38;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 9 LLMGYTWC 16
:|:|:|:|:|
Db 109 LVMGWSWG 116

RESULT 20
T04608
ADP, ATP carrier protein F2009.60 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 21-May-1999
C:Accession: T04608
R:Boyan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hohelsel, J.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, October 1998
A:Reference number: 215380
A:Accession: T04608
A:Molecule type: DNA
A:Residues: 1-379 <BREV>
A:Cross-references: EMBL:AL021749
A:Experimental source: cultivar Columbia; BAC clone F2009
C:Genetics:
A:Map position: 4
A:Introns: 157/3; 281/3
A:Note: F2009.60
C:Superfamily: ADP, ATP carrier protein; ADP, ATP carrier protein repeat homology
C:Keywords: duplication; transmembrane protein
F:76-171/Domain: ADP, ATP carrier protein repeat homology <ACPI>
F:181-275/Domain: ADP, ATP carrier protein repeat homology <ACP2>
F:261-369/Domain: ADP, ATP carrier protein repeat homology <ACP3>

Query Match 50.6%; Score 44; DB 2; Length 379;
Best Local Similarity 60.0%; Pred. No. 43;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 SMLVYFLLMGVTWC 16
| | | | | | | | | |
Db 283 SFLASFLMCITIG 297

RESULT 21
H83708
Hypotheetical protein BH0472 [Imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: H83708
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H.
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20263314
A:Accession: H83708
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-417 <STO>
A:Cross-references: GB:AP001508; GB:BA000004; NID:q10172890; PIDN:BAH04191.1; GSPDB:C
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0472

Query Match 50.6%; Score 44; DB 2; Length 417;
Best Local Similarity 54.5%; Pred. No. 46;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 VVFLLMGVTV 15
:|:|:|:|:|
Db 104 VILRLMGJAW 114

RESULT 22
T49017
Hypotheetical protein F3C22.40 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T49017
R:Purnelle, B.; Masuy, D.; Goffeau, A.; Boutry, M.; Mewes, H.W.; Rudd, S.; Lemcke, K.
submitted to the Protein Sequence Database, April 2000
A:Reference number: 225013
A:Accession: T49017
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <PUR>
A:Cross-references: EMBL:AL353912; GSPDB:GN00061; ATSP:F3C22.40
A:Experimental source: cultivar Columbia; BAC clone F3C22
C:Genetics:
A:Gene: ATSP:F3C22.40
A:Map position: 3
A:Introns: 18/3; 43/2; 87/3; 111/3; 151/3; 175/3; 222/3; 251/1; 399/2
C:Superfamily: Arabidopsis thaliana hypotheetical protein F3C22.40

Query Match 50.6%; Score 44; DB 2; Length 486;
Best Local Similarity 66.7%; Pred. No. 52;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 5 VVFLLMGVTV 16
:|:|:|:|:|
Db 112 LVFLVLTGERTWC 123

RESULT 23
H96681
Protein F1E22.10 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H96681
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hutzler, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Lin, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H96681
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-688 <STO>
A:Cross-references: GB:AE005173; NID:9666396; PIDN:AAF23030.1; GSPDB:GN00141
C:Genetics:
A:Gene: F1E22.10
A:Map position: 1

Query Match 50.6%; Score 44; DB 2; Length 688;
Best Local Similarity 46.7%; Pred. No. 68;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Oy 1 MSMLVFLLMGVTW 15
Db 281 VSLIGALILMGVNM 295

RESULT 24

hypothetical protein XF1330 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: DB2696
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: For a complete list of authors see reference number A59328 below
A:Accession: DB2696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1197 <STM>
A:Cross-references: GB:AE003965; GB:AE003849; NID:99106313; PIDN:AAF4139.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
B:Jones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facinanci, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigt
Chado, M.A.; Medeira, A.M.B.N.; Medeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, B
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiti, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1330

Query Match 50.6%; Score 44; DB 2; Length 1197;
Best Local Similarity 87.5%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 8 LLLMGVTW 15
Db 771 LLLMGVAM 778

RESULT 25
conserved hypothetical protein yvbv - Bacillus subtilis
F70030

C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: F70030
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C:Bro, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Caplan, V.; Carter, N.M.;
A:Enrich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Gal
leach, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpiltra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danochin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033
A:Accession: F70030
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-305 <KUN>
A:Cross-references: GB:Z29121; GB:AL009126; NID:92635827; PIDN:CAB15405.1; PID:el1860
A:Experimental source: strain 168
C:Genetics:
A:Gene: yvbv

Query Match 50.0%; Score 43.5; DB 2; Length 305;
Best Local Similarity 46.7%; Pred. No. 43;
Matches 7; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

Oy 2 SMLVFL-LLMGVTW 15
Db 10 ALLAFVLMGVNM 24

RESULT 26

hypothetical protein slr0881 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S77270
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S77270
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-103 <KAN>
A:Cross-references: EMBL:D90907; GB:AB001339; NID:91652618; PIDN:BAAL1604.1; PID:9165
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Synechocystis hypothetical protein slr0881

Query Match 49.4%; Score 43; DB 2; Length 103;
Best Local Similarity 33.3%; Pred. No. 22;
Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Oy 1 MSMLVFLLMGVTW 15
Db 2 LALLAIALLMGLIM 16

RESULT 27

F39925
hypothetical protein 5 - equine arteritis virus
C:Species: equine arteritis virus
C:Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 08-Oct-1999
C:Accession: F39925
R:Don Boon, J.A.; Snijder, E.J.; Chirnside, E.D.; De Vries, A.A.F.; Horzinek, M.C.; Spaa
J. Virol. 65, 2910-2920, 1991
A:Title: Equine arteritis virus is not a togavirus but belongs to the coronavirusslike su
A:Reference number: A39925; MUID:91237805
A:Accession: F39925
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-255 <DEN>
A:Cross-references: EMBL:X53459; NID:g62065; PIDN:CAA37544.1; PID:g62071

Query Match 49.4%; Score 43; DB 1; Length 255;
Best Local Similarity 58.3%; Pred. No. 44;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 MSMLVFLLMG 12
:||:||||
Db 2 LSMIVLFLMG 13

RESULT 28
S52775
hypothetical protein 2 - Chloroflexus aurantiacus
C:Species: Chloroflexus aurantiacus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S52775
R:Niedemeyer, G.; Shiozawa, J.A.; Lottspeich, F.; Felck, R.G.
FEBS Lett. 342, 61-65, 1994
A:Title: The primary structure of two chlorosome proteins from Chloroflexus aurantiacus.
A:Reference number: S43678; MUID:94192803
A:Accession: S52775
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <NIE>
A:Cross-references: EMBL:Z34000; NID:g496485; PIDN:CAA63969.1; PID:g496488
A:Note: only a part of the coding sequence is given in this paper
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994
C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0279

Query Match 49.4%; Score 43; DB 1; Length 305;
Best Local Similarity 42.9%; Pred. No. 50;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Oy 2 SMLVFLMGVTW 15
:||:|||||
Db 237 AMLMILALMGQYW 250

RESULT 29
S30259
ADP/ATP carrier protein, mitochondrial - Chlamydomonas reinhardtii
N:Alternate names: ADP/ATP translocator
C:Species: Chlamydomonas reinhardtii
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S30259; S20628
R:Sharpe, J.A.; Day, A.
Mol. Gen. Genet. 237, 134-144, 1993
A:Title: Structure, evolution and expression of the mitochondrial ADP/ATP translocator g
A:Reference number: S30259; MUID:95204887
A:Accession: S30259
A:Molecule type: mRNA
A:Residues: 1-308 <SHA>
A:Cross-references: EMBL:X65194; NID:g18109; PIDN:CAA6311.1; PID:g18110
C:Superfamily: ADP/ATP carrier protein; ADP/ATP carrier protein repeat homology
C:Keywords: duplication; homodimer; mitochondrial; transmembrane protein
F:5-100/Domain: ADP/ATP carrier protein repeat homology <ACPI>
F:109-204/Domain: ADP/ATP carrier protein repeat homology <ACP2>

F:210-298/Domain: ADP/ATP carrier protein repeat homology <ACP3>

Query Match 49.4%; Score 43; DB 1; Length 308;
Best Local Similarity 53.3%; Pred. No. 50;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 2 SMLVFLMGVTWC 16
:|||||||
Db 212 NFLAFLLMGVITIG 226

RESULT 30
S51791
Drosophila translocation protein 1 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
C:Accession: S51791
R:Noel, P.J.; Cartwright, I.L.
EMBO J. 13, 5253-5261, 1994
A:Title: A Sec62p-related component of the secretory protein translocon from Drosophi
A:Reference number: S51791; MUID:95043419
A:Accession: S51791
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-402 <NOE>
A:Cross-references: EMBL:Z38100; NID:g558180; PIDN:CAA66222.1; PID:g558181
C:Genetics:
A:Gene: FlyBase:Trp1
A:Gene: FlyBase:FBgn0011584
A:Cross-references: FlyBase:FBgn0011584
C:Superfamily: Caenorhabditis elegans hypothetical protein C18E9.2

Query Match 49.4%; Score 43; DB 2; Length 402;
Best Local Similarity 46.2%; Pred. No. 62;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 4 LVVFLMGVTWC 16
:|||||||
Db 247 LVVFLMGVTWC 259

RESULT 31
S29483
rhodopsin [similarity] - Japanese flying squid
N:Alternate names: visual pigment protein
C:Species: Todarodes pacificus (Japanese flying squid)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C:Accession: S29483; PT0063
R:Hara-Nishimura, I.; Kondo, M.; Nishimura, M.; Hara, R.; Hara, T.
FEBS Lett. 317, 5-11, 1993
A:Title: Cloning and nucleotide sequence of cDNA for rhodopsin of the squid Todarodes
A:Reference number: S29483; MUID:93154520
A:Accession: S29483
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-448 <HAR>
A:Cross-references: EMBL:X70498; NID:g397389; PIDN:CAA9906.1; PID:g397390
R:Seidou, M.; Kubota, I.; Hiraki, R.; Kito, Y.
Biochim. Biophys. Acta 957, 318-321, 1998
A:Title: Amino acid sequence of the retinal binding site of squid visual pigment.
A:Reference number: PT0063; MUID:89051045
A:Accession: PT0063
A:Molecule type: protein
A:Residues: 303-313 <SEI>
C:Superfamily: octopus rhodopsin
F:36-60/Domain: transmembrane #status predicted <TM1>
F:73-97/Domain: transmembrane #status predicted <TM2>
F:106-130/Domain: transmembrane #status predicted <TM3>
F:152-176/Domain: transmembrane #status predicted <TM4>
F:200-223/Domain: transmembrane #status predicted <TM5>
F:262-285/Domain: transmembrane #status predicted <TM6>

F:301-322/Domain: transmembrane #status predicted <TM>
F:8.14/Binding site: carboxylate (Asn) (covalent) #status predicted
F:305/Binding site: retinal (Lys) (covalent) #status experimental
F:336,337/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 49.4% Score 43; DB 1; Length 448;
Best Local Similarity 38.5% Pred. No. 67;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 3 MLVFLILMGVTW 15
: : : : :
Db 154 IMIFVWLMSVLM 166

RESULT 32
H84922
hypothetical protein At2g48060 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84922
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487
A:Accession: H84922
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-621 <STO>
A:Cross-references: GB:AE002093; NID:g4249417; PIDN:ABD13714.1; GSPDB:GND0139
C:Genetics:
A:Gene: At2g48060
A:Map position: 2

Query Match 49.4% Score 43; DB 2; Length 621;
Best Local Similarity 58.3% Pred. No. 86;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 3 MLVFLILMGVT 14
: : : : :
Db 405 LLVFLILMAVS 416

RESULT 33
T05361
hypothetical protein F8B4.210 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05361
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysheart, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, February 1999
A:Reference number: 215409
A:Accession: T05361
A:Molecule type: DNA
A:Residues: 1-710 <BEV>
A:Cross-references: EMBL:AL034567
A:Experimental source: cultivar Columbia; BAC clone F8B4
C:Genetics:
A:Map position: 4
A:Introns: 44/2; 75/2; 137/2; 190/3; 246/1; 301/3; 359/3; 391/3; 450/3; 559/2
A>Note: F8B4.210

Query Match 49.4% Score 43; DB 2; Length 710;
Best Local Similarity 58.3% Pred. No. 95;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
OY 4 LVVFLILMGVTW 15
: : : : :
1: 11: 1111

Db 585 LLVFLILMGVTW 596

RESULT 34
T13423
hypothetical protein 30B8.4 - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13423
R:Murphy, L.; Harris, D.; Barrell, B.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: 217668
A:Accession: T13423
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-3345 <MDR>
A:Cross-references: EMBL:AL009195; NID:e1355203; PID:e1248585; PIDN:CAA15708.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0000377
A:Map position: X
A:Introns: 51/3; 159/1; 476/1; 526/1; 1465/1; 1826/3; 1947/3; 2081/1; 2196/3; 3007/3
A>Note: EG:30B8.4

Query Match 49.4% Score 43; DB 2; Length 3345;
Best Local Similarity 31.2% Pred. No. 31e+02;
Matches 5; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 1 MSMLVFLILMGVTWG 16
: : : : :
Db 1948 LQFIVTYIAPMQITWG 1963

RESULT 35
F81111
nickel-dependent hydrogenase, b-type cytochrome chain NMB1158 [imported] - N
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: F81111; F81115
R:Retlein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
rl, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: F81111
A:Molecule type: DNA
A:Residues: 1-218 <TEP>
A:Cross-references: GB:AE002467; GB:AE002098; NID:g7226426; PIDN:AAF41579.1; PID:g722
A:Experimental source: serogroup B, strain MC58
A:Accession: F81115
A:Molecule type: DNA
A:Residues: 1-218 <TEZ>
A:Cross-references: GB:AE002464; GB:AE002098; NID:g7226388; PIDN:AAF41544.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1158

Query Match 48.9% Score 42.5; DB 2; Length 218;
Best Local Similarity 61.5% Pred. No. 45;
Matches 8; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

OY 4 LVVFLILMGVTWG 16
: : : : :
Db 50 LNVFLILCMGI-WG 61

RESULT 36
B81906
probable membrane protein NMA1368 [imported] - Neisseria meningitidis (strain 22491 s

C:Species: *Neisseria meningitidis*
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: AB1906
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A:Reference number: AB1775; MUID:20222556
A:Accession: AB1906
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-218 <PAR>
A:Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CA884614.1; PID:g738003
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA1368

Query Match 48.9%; Score 42.5; DB 2; Length 218;
Best Local Similarity 61.5%; Pred. No. 45;
Matches 8; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

OY 4 LVFLLMGVTWG 16
||| | | | | |
Db 50 LVFRLCWGI-WG 61

RESULT 37
H25035
hypothetical protein 2 - *Escherichia coli* plasmid COL1b
C:Species: *Escherichia coli*
C:Date: 24-Jan-1988 #sequence_revision 24-Jan-1988 #text_change 18-Jun-1993
C:Accession: H25035
R:Mankovich, J.A.; Hsu, C.H.; Konisky, J.
J. Bacteriol. 168, 228-236, 1986
A:Title: DNA and amino acid sequence analysis of structural and immunity genes of colic
A:Reference number: A91822; MUID:87008385
A:Accession: H25035
A:Molecule type: DNA
A:Residues: 1-102 <MAN>
C:Genetics:
A:Genome: plasmid

Query Match 48.3%; Score 42; DB 2; Length 102;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 10 LMGVTVG 16
||| | | | | |
Db 44 LMGISWG 50

RESULT 38
G25035
hypothetical protein 2 - *Escherichia coli* plasmid COL1a
C:Species: *Escherichia coli*
C:Date: 24-Jan-1988 #sequence_revision 24-Jan-1988 #text_change 18-Jun-1993
C:Accession: G25035
R:Mankovich, J.A.; Hsu, C.H.; Konisky, J.
J. Bacteriol. 168, 228-236, 1986
A:Title: DNA and amino acid sequence analysis of structural and immunity genes of colic
A:Reference number: A91822; MUID:87008385
A:Accession: G25035
A:Molecule type: DNA
A:Residues: 1-102 <MAN>
C:Genetics:
A:Genome: plasmid

Query Match 48.3%; Score 42; DB 2; Length 102;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 10 LMGVTVG 16
||| | | | | |
Db 44 LMGISWG 50

RESULT 39
E72636
probable modulation ATP-binding protein I APE1548 - *Aeropyrum pernix* (strain K1)
C:Species: *Aeropyrum pernix*
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 17-Mar-2000
C:Accession: E72636
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic *Crenarchaeon*, *Aero*
A:Reference number: A72450; MUID:99310339
A:Accession: E72636
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <KAM>
A:Cross-references: DBJ:AP000061; NID:g5104821; PIDN:BA80547.1; PID:d1044333; PID:g
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1548
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match 48.3%; Score 42; DB 2; Length 338;
Best Local Similarity 57.1%; Pred. No. 74;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 3 MLVFLLMGVTVG 16
|| | | | | | |
Db 25 MGVGPVWGVWVG 38

RESULT 40
AB3795
hypothetical protein BH1161 [Imported] - *Bacillus halodurans* (strain C-125)
C:Species: *Bacillus halodurans*
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: AB3795
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a
A:Reference number: AB3650; MUID:20263314
A:Accession: AB3795
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-406 <STO>
A:Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BA804880.1; GSPDB:C
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1161

Query Match 48.3%; Score 42; DB 2; Length 406;
Best Local Similarity 50.0%; Pred. No. 86;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 4 LVFLLMGVTWG 15
|| | | | | | |
Db 97 LIIARILMGVAM 108

RESULT 41
T36979
probable integral membrane transport protein - *Streptomyces coelicolor*
C:Species: *Streptomyces coelicolor*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T36979
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A

submitted to the EMBL Data Library, August 1999
A:Reference number: 221618

A:Accession: T36979

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-407 <O11>

A:Cross-references: EMBL:AL109949; PIDN:GAB52893.1; GSPDB:GN00070; SCOPDB:SCJ11.08C

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCJ11.08C

C:Superfamily: Streptomyces lividans chloramphenicol resistance protein

Query Match 48.3%; Score 42; DB 1; Length 407;
Best Local Similarity 58.3%; Pred. No. 86;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 VVFLLMGVTWG 16
DB 313 VALLVMGAVG 324

RESULT 42
S24065
preprotein translocase secY [validated] - Methanococcus vannielii

C:Species: Methanococcus vannielii

C>Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 18-Feb-2000

C:Accession: S24065

R:Auer, J.; Spicker, G.; Boeck, A.

Biochimie 73, 683-688, 1991

A:Title: Presence of a gene in the archaeobacterium Methanococcus vannielii homologous to

A:Reference number: S24065; MUID:92110434

A:Accession: S24065

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-438 <AUE>

A:Cross-references: EMBL:X62045; NID:g44771; PIDN:CAAA3978.1; PID:g44772

C:Function:

A:Description: essential for preprotein translocation across the cytoplasmic membrane [y

A>Note: Methanococcus secY protein can complement a ts mutant in Escherichia coli

C:Superfamily: yeast SSH1 protein

C:Keywords: protein transport; transmembrane protein

F:30-46/Domain: transmembrane #status predicted <TM1>

F:72-88/Domain: transmembrane #status predicted <TM2>

F:112-128/Domain: transmembrane #status predicted <TM3>

F:138-154/Domain: transmembrane #status predicted <TM4>

F:169-185/Domain: transmembrane #status predicted <TM5>

F:211-227/Domain: transmembrane #status predicted <TM6>

F:252-268/Domain: transmembrane #status predicted <TM7>

F:318-334/Domain: transmembrane #status predicted <TM8>

F:377-393/Domain: transmembrane #status predicted <TM9>

Query Match 48.3%; Score 42; DB 1; Length 438;
Best Local Similarity 46.2%; Pred. No. 91;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 3 MLVFLLMGVTW 15
DB 323 MVVFCILGFLTW 335

RESULT 43
O00CG

Rhodopsin - giant octopus

C:Species: Octopus doylei (giant octopus)

C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 05-Jun-1998

C:Accession: S00610

R:Ovchinnikov, Y.A.; Abdulaev, N.G.; Zolotarev, A.S.; Artamonov, I.D.; Bepalov, I.A.; D

FEBS Lett. 232, 69-72, 1988

A:Title: Octopus rhodopsin. Amino acid sequence deduced from cDNA.

A:Reference number: S00610; MUID:88211878

A:Accession: S00610

A:Molecule type: mRNA

A:Residues: 1-455 <OVC>

A:Cross-references: EMBL:X07797

A>Note: the source is designated as Paroctopus defleini

A:Note: part of this sequence was confirmed by protein sequencing

C:Superfamily: octopus rhodopsin

C:Keywords: chromoprotein; color vision; G protein-coupled receptor; glycoprotein; 11

F:37-61/Domain: transmembrane #status predicted <TM1>

F:74-98/Domain: transmembrane #status predicted <TM2>

F:107-131/Domain: transmembrane #status predicted <TM3>

F:153-177/Domain: transmembrane #status predicted <TM4>

F:201-224/Domain: transmembrane #status predicted <TM5>

F:263-286/Domain: transmembrane #status predicted <TM6>

F:302-323/Domain: transmembrane #status predicted <TM7>

F:9.15/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:306/Binding site: retinal (Lys) (covalent) #status predicted

F:337/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 48.3%; Score 42; DB 1; Length 455;
Best Local Similarity 23.1%; Pred. No. 93;
Matches 3; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

OY 3 MLVFLLMGVTW 15
DB 155 LMIFVMMSIWM 167

RESULT 44
S28306
hypochemical protein T2365.5 - Caenorhabditis elegans (fragment)

C:Species: Caenorhabditis elegans

C>Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 21-Aug-1998

C:Accession: S28306

R:Berk, M.

submitted to the EMBL Data Library, December 1992

A:Reference number: S28296

A:Accession: S28306

A:Molecule type: DNA

A:Residues: 1-499 <BER>

A:Cross-references: EMBL:Z19158

C:Genetics:

A:Insertion: 27/3; 134/2; 184/2; 215/3; 253/2; 332/3; 375/3; 409/1; 442/3

C:Superfamily: gamma-aminobutyric acid transporter

Query Match 48.3%; Score 42; DB 2; Length 499;
Best Local Similarity 35.7%; Pred. No. 1e+02;
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

OY 3 MLVFLLMGVTWG 16
DB 379 LIMEWLIYGTWG 392

RESULT 45
G64902

extreme acid resistance protein xasa - Escherichia coli

N:Alternate names: probable permease xasa

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2000

C:Accession: G64902; S78627

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: G64902

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-511 <BLAT>

A:Cross-references: GB:AE000246; GB:U00096; NID:g1787764; PIDN:AACT4565.1; PID:g17877

A:Experimental source: strain K-12, substrain MG1655

R:Herst, B.M.; Farooq, F.T.; Barstad, D.N.; Blankenhorn, D.L.; Sionczewski, J.L.
J. Bacteriol. 178, 3978-3981, 1996
A:Title: A glutamate-dependent acid resistance gene in *Escherichia coli*.
A:Reference number: S78627; MUID:96272279
A:Accession: S78627
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 402-489 <HER>
A:Cross-references: EMBL:U13204
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
C:Genetics:
A:Gene: xasa; gadC
C:Function:
A:Description: may act as a transporter for the product of glutamate decarboxylase, gamma
A:Note: part of a glutamate decarboxylase alkalization cycle to protect *E. coli* from c
C:Superfamily: arginine permease
C:Keywords: amino acid transport; inner membrane; transmembrane protein
F:13-29/Domain: transmembrane #status predicted <TM1>
F:42-58/Domain: transmembrane #status predicted <TM2>
F:99-115/Domain: transmembrane #status predicted <TM3>
F:128-144/Domain: transmembrane #status predicted <TM4>
F:161-177/Domain: transmembrane #status predicted <TM5>
F:201-217/Domain: transmembrane #status predicted <TM6>
F:236-252/Domain: transmembrane #status predicted <TM7>
F:290-306/Domain: transmembrane #status predicted <TM8>
F:339-355/Domain: transmembrane #status predicted <TM9>
F:371-387/Domain: transmembrane #status predicted <TM10>
F:413-429/Domain: transmembrane #status predicted <TM11>
F:448-464/Domain: transmembrane #status predicted <TM12>

Query Match 48.3%; Score 42; DB 2; Length 511;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 4 LVFPLLLNGVTW 15
||| ||| | : |
Db 42 LVFPLLLGGLW 53

Search completed: January 7, 2002, 16:52:10
Job time: 227 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2002, 16:50:33 ; Search time 16.52 Seconds

(Without alignments)
35.511 Million cell updates/sec

Title: US-09-471-276-831_COPY_1_16

Perfect score: 87

Sequence: 1 MSMLVFFLLMGVTWG 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%

Listing first 100 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	49	56.3	217	1 YPRA_ECOLI	P13974 escherichia
2	49	56.3	1107	1 YJEP_ECOLI	P39285 escherichia
3	47	54.0	48	1 YC21_ARCFU	O29047 archaeoglob
4	46	52.9	374	1 Y006_BORBU	O51039 borrelia bu
5	44	50.6	323	1 PIP-STRCO	O98214 streptomyce
6	43.5	50.0	305	1 YVBV_BACSU	O32256 bacillus su
7	43	49.4	255	1 YOR5_EAV	P28995 equine arte
8	43	49.4	308	1 ADT_CHLRE	P27080 chlamydomon
9	43	49.4	447	1 OPSD_TOLPA	P31356 todarodes p
10	42.5	48.9	304	1 PAGO_SALTY	O30646 salmonella
11	42	48.3	438	1 SECY_MERYA	P28541 methanococ
12	42	48.3	455	1 OPSD_OCTDO	P09241 octopus dof
13	42	48.3	511	1 XASA_ECO57	P39183 escherichia
14	42	48.3	511	1 XASA_ECOLI	O54152 shigella fl
15	42	48.3	543	1 XASA_SHIFL	O54152 shigella fl
16	42	48.3	543	1 DAL5_YEAST	P15365 saccharomyc
17	42	48.3	615	1 NTPD_CAREL	O03614 caenorhadi
18	41	47.1	241	1 MXIU_SHIFL	O08081 shigella fl
19	41	47.1	241	1 MXIU_SHISO	O08081 shigella so
20	41	47.1	301	1 YJJE_ECOLI	O52288 shigella so
21	41	47.1	353	1 OPSD_LIZSA	O9Y927 liza salien
22	41	47.1	353	1 OPSD_MUGCE	O9Y929 mugil cepha
23	41	47.1	546	1 NUSM_CHLRE	P08739 chlamydomon
24	40.5	46.6	797	1 ATIS_HUMAN	P09196 homo sapien
25	40	46.0	141	1 USHR_TRIUV	O46488 trichosurus
26	40	46.0	211	1 YFHB_ECOLI	P30132 escherichia
27	40	46.0	259	1 RPEI_KLEPN	O48475 klebsiella
28	40	46.0	278	1 RCEL_CHRYI	P51762 chromatiu
29	40	46.0	304	1 OPSD_ICTPU	O42268 ictalurus p
30	40	46.0	305	1 ADT_KLUUY	P49382 kluyveromyc
31	40	46.0	319	1 CHIB_SALTY	O05600 salmonella
32	40	46.0	349	1 YHHT_ECOLI	P37622 escherichia
33	40	46.0	353	1 OPSD_CHELB	O9Y928 chelon labr

34	40	46.0	353	1 OPSD_LITMO	O9Y900 lithognathu
35	40	46.0	353	1 OPSD_LIZAU	O9Y926 liza aurata
36	40	46.0	353	1 OPSD_SOLOU	O9Y925 soloa soloa
37	40	46.0	354	1 OPSU_BRARE	P35359 brachydanio
38	40	46.0	446	1 GUDX_ECOLI	O46915 escherichia
39	40	46.0	451	1 GUDH_PSEPU	P42206 pseudomonas
40	40	46.0	455	1 ILXN_BACSU	P42238 bacillus su
41	40	46.0	660	1 FHUB_ECOLI	P06972 escherichia
42	39.5	45.4	405	1 Y281_BHCAI	P57368 buchiera ap
43	39.5	45.4	455	1 UHPT_CHLNP	O92719 chlamydia p
44	39.5	45.4	456	1 UHPT_CHLTR	O84548 chlamydia t
45	39	44.8	157	1 YXJN_BACSU	P55182 bacillus su
46	39	44.8	268	1 ILXA_MOUSE	P01590 mus musculu
47	39	44.8	277	1 TSG6_HUMAN	P98066 homo sapien
48	39	44.8	292	1 VOAV_BACSU	O34416 bacillus su
49	39	44.8	341	1 CATL_DROME	O35029 drosophila
50	39	44.8	342	1 Y762_MERJA	O58172 methanococc
51	39	44.8	353	1 OPSD_SPANU	O9Y902 sparus aura
52	39	44.8	360	1 PCS2_PIG	O9X549 sus scrofa
53	39	44.8	387	1 HM74_HUMAN	P49019 homo sapien
54	39	44.8	450	1 OSTA_SCHPO	O10176 schizosach
55	39	44.8	522	1 STA_RITCO	O10710 rictinus com
56	39	44.8	910	1 DDRI_MOUSE	O63474 ratus norv
57	39	44.8	911	1 DDRI_MOUSE	O03146 mus musculu
58	39	44.8	913	1 DDRI_MOUSE	O08345 homo sapien
59	39	44.8	1321	1 AB11_RABIT	O9N093 oryctolagus
60	38.5	44.3	349	1 I10S_MOUSE	O61199 mus musculu
61	38.5	44.3	456	1 UHPT_CHLMU	O9Y918 chlamydia m
62	38	43.7	43	1 VE5_PAPYE	P11330 european el
63	38	43.7	169	1 X192_ECOLI	P17738 escherichia
64	38	43.7	198	1 YB0J_MYCPN	P75806 escherichia
65	38	43.7	251	1 YB0J_MYCPN	P75566 mycoplasma
66	38	43.7	272	1 IL2A_HUMAN	P01589 homo sapien
67	38	43.7	322	1 N01M_PELSU	O79670 pelionedusa
68	38	43.7	329	1 CY51_CABEL	P25807 caenorhadi
69	38	43.7	335	1 CD1D_HUMAN	P15813 homo sapien
70	38	43.7	360	1 PCS2_CANPA	O29393 canis fami
71	38	43.7	360	1 PCS2_HORSE	O46542 equus cabal
72	38	43.7	375	1 GPRS_HUMAN	O9N567 homo sapien
73	38	43.7	377	1 GPRS_RAT	O9Y918 ratus norv
74	38	43.7	379	1 GPRS_MOUSE	O9Y918 ratus norv
75	38	43.7	402	1 OPDE_PSEAE	O01602 pseudomonas
76	38	43.7	416	1 TUTB_ERHME	O47825 erwina her
77	38	43.7	423	1 IRK6_HUMAN	P48051 homo sapien
78	38	43.7	425	1 IRK6_MESNU	P49658 mesocricetu
79	38	43.7	425	1 IRK6_MOUSE	P48542 mus musculu
80	38	43.7	425	1 IRK6_MOUSE	P48542 mus musculu
81	38	43.7	439	1 OPSD_LOLSU	O47094 loligo subu
82	38	43.7	449	1 YGIV_ECOLI	O17094 escherichia
83	38	43.7	464	1 OPSD_SEPOF	O16005 sepiia offic
84	38	43.7	465	1 MOT4_HUMAN	O15427 homo sapien
85	38	43.7	470	1 MOT4_MOUSE	P57787 mus musculu
86	38	43.7	471	1 MOT4_MOUSE	O35910 ratus norv
87	38	43.7	471	1 Y872_HAEIN	O57491 haemophilus
88	38	43.7	473	1 MOT4_CHICK	P57788 gallus galli
89	38	43.7	480	1 PRSB_STAXY	P51184 straphylococ
90	38	43.7	563	1 CYB_SULAC	P39480 sulfolobus
91	38	43.7	614	1 YDNK_LACLC	P42377 lactococcus
92	38	43.7	757	1 SPT3_CAEEL	P46973 caenorhadi
93	38	43.7	799	1 ITBN_DROME	O02521 drosophila
94	38	43.7	837	1 YP84_CAEEL	O09725 caenorhadi
95	38	43.7	984	1 SX13_MOUSE	O04891 mus musculu
96	38	43.7	1676	1 COS_HUMAN	O01031 homo sapien
97	38	43.7	1742	1 GUNA_CALSA	P22534 caldocellum
98	37.5	43.1	261	1 KUK3_MACMU	P33616 macaca mula
99	37.5	43.1	271	1 COMT_HUMAN	P21964 homo sapien
100	37.5	43.1	282	1 VMSA_MHYW6	P11293 woodchuck h

ALIGNMENTS

RESULT 1

YPR_A_ECOLI STANDARD; PRT: 217 AA.
AC P13974;
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHEETICAL 24.3 KDA PROTEIN (URF 1).
OS Escherichia coli.
OC Plasmid IncFII R1.
CC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID:562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:88121694; PubMed:3323833;
RA Bravo A., de Torreotequi G., Diaz R.;
RT "Identification of components of a new stability system of plasmid
RT R1, PRD, that is close to the origin of replication of this
RT plasmid.";
RL Mol. Gen. Genet. 210:101-110(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X06240; CAA29583.1; -;
DR PIR: S01095; S01095.
DR InterPro: IPR003675; ABL.
DR Pfam: PF02517; ABL.1.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 217 AA; 24307 MW; EA519B91C43B6666 CRC64;

Query Match 56.3%; Score 49; DB 1; Length 217;
Best Local Similarity 37.5%; Pred. No. 3.9;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSMLVFLILMGVTWG 16
Db 9 LQFMLVLAACISWG 24

RESULT 2
YPR_A_ECOLI STANDARD; PRT: 1107 AA.
ID YPR_A_ECOLI
AC P39285; P76798;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEETICAL PROTEIN YPR PRECURSOR.
OS YPR OR B4159.
GN Escherichia coli.
OC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID:562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN:K12 / MG1655;
RA MEDLINE:9534362; PubMed:7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP REVISION TO 1015.
RC STRAIN:K12 / MG1655;
RX MEDLINE:9742617; PubMed:9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN:K12;
RX MEDLINE:88298809; PubMed:3042771;
RA Li O.X., Downham W.;
RT "Structural characterization of Escherichia coli phosphatidylinositol
RT decarboxylase.";
RL J. Biol. Chem. 263:11516-11522(1988).
CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1 SIMILARITY: BELONGS TO THE UPF0003 FAMILY. STRONG, TO H. INFLUENZAE
CC H10195.1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14003; AAA97058.1; -;
DR EMBL: AE000488; AAC77119.1; -;
DR EMBL: J03916; AAA83897.1; ALT_INIT.
DR EcoGene: EG12478; yjeP.
DR InterPro: IPR001880; UPF00003.
DR Pfam: PF00924; UPF0003; 1.
DR PROSITE: PS01246; UPF0003; 1.
KW Hypothetical protein; Transmembrane; Signal; Complete proteome.
FT SIGNAL 1 19
FT CHAIN 20 1107
FT TRANSMEM 467 487
FT TRANSMEM 551 571
FT TRANSMEM 600 620
FT TRANSMEM 628 648
FT TRANSMEM 674 694
FT TRANSMEM 698 718
FT TRANSMEM 785 805
FT TRANSMEM 828 848
FT TRANSMEM 875 895
FT TRANSMEM 910 930
FT CONFLICT 1015 1015 R -> A (IN REF. 1).
SQ SEQUENCE 1107 AA; 123967 MW; 5F52A293B9052B CRC64;

Query Match 56.3%; Score 49; DB 1; Length 1107;
Best Local Similarity 37.5%; Pred. No. 14;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSMLVFLILMGVTWG 16
Db 1 MRLIITFLMAMCLSWG 16

RESULT 3
YC21_ARCFU STANDARD; PRT: 48 AA.
ID YC21_ARCFU
AC O29047;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEETICAL PROTEIN AF1221.
GN AF1221.
OS Archaeoglobus fulgidus.
OC Archaea: Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus
OX NCBI_TaxID:2234;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RA MEDLINE-98049343; PubMed-9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Kirschner K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kiehl J.-F., Fleischmann A.R., Kyrle S.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.T., McNeil L.K., Badger J.H., Glöckel A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spillars T., Arltach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon *Archaeoglobus fulgidus*.";
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001020; AAB90036.1; -
DR TIGR: AF1221; -
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 25 47 POTENTIAL.
SQ SEQUENCE 48 AA: 5307 MW: 6AAB25353BB27BC1 CRC64:

Query Match 54.0%; Score 47; DB 1; Length 48;
Best Local Similarity 54.5%; Pred. No. 2.2;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 5 VFFLLMGVTW 15
Db 37 IVILLMGISM 47

RESULT 4
Y006_BORBU STANDARD: PRT; 374 AA.
AC 051039;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN BB0006.
GN BB0006.
OS *Borrelia burgdorferi* (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; *Borrelia*.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35210 / B31;
RX MEDLINE-98065943; PubMed-9403685;
RA Fraser C.M., Chasens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Iachira R., White O., Kechum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kiehl J.-F., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Matthey L., McDonald L., Arltach P., Bowman C.,
RA Smith H.O., Venter J.C.;
RA "Genomic sequence of a Lyme disease spirochete, *Borrelia*
RT *burgdorferi*.";
RT Nature 390:580-586(1997).
RL -----
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE UPF0118 (PERM) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001115; AAC66397.1; -
DR TIGR: BB0006; -
DR InterPro: IPR002549; UPF0118.
DR Pfam: PF01594; UPF0118; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 32 52 POTENTIAL.
FT TRANSMEM 55 75 POTENTIAL.
FT TRANSMEM 80 100 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 224 244 POTENTIAL.
FT TRANSMEM 246 266 POTENTIAL.
FT TRANSMEM 310 330 POTENTIAL.
FT TRANSMEM 336 356 POTENTIAL.
SQ SEQUENCE 374 AA: 42910 MW: 905948C4B0646900 CRC64:

Query Match 52.9%; Score 46; DB 1; Length 374;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 MLVFFLLMGVTWG 16
Db 311 LILCFLLPFGWIMWG 324

RESULT 5
PIP_STRCO STANDARD: PRT; 323 AA.
AC Q9S2L4;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PROLINE IMINOPEPTIDASE (EC 3.4.11.5) (PIP) (PROLYL
DE AMINOPEPTIDASE) (PAP).
GN SC7H2.03C.
GN Streptomyces coelicolor.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Saunders D.C., Harris D., Bentley S.D., Parkhill J., Barrell B.G.,
RA Rajandream M.A.; 1999) to the EMBL/GenBank/DBJ databases.
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: SPECIFICALLY CATALYZES THE REMOVAL OF N-TERMINAL PROLINE
CC RESIDUES FROM PEPTIDES. (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: RELEASE OF A N-TERMINAL PROLINE FROM A
CC PEPTIDE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S33.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL109732; CAB52045.1; -
DR InterPro: IPR000073; Abhydrolase.
DR InterPro: IPR000379; Est_11p_Thioestl_actsite.
DR InterPro: IPR002410; Pro_aminopeptase.
DR Pfam: PF00561; abhydrolase; 1.
DR PRINTS: PR00793; PROAMNPTASE.
KW Hydrolase; Aminopeptidase.

FT ACT_SITE 114 114 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 271 271 BY SIMILARITY.
 FT ACT_SITE 299 299 PROTON DONOR (BY SIMILARITY).
 SQ SEQUENCE 323 AA; 35040 MW; 7BEF00133B353EEB CRC64;

Query Match 50.6%; Score 44; DB 1; Length 323;
 Best Local Similarity 75.0%; Pred. No. 25;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 9 LLMGVTVG 16
 Db 109 LLMGVTVG 116

RESULT 6
 YVBY_BACSU STANDARD; PRT; 305 AA.
 AC O32256;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 32.8 KDA PROTEIN IN ARAR-LACA INTERGENIC REGION.
 GN YVBY.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID:1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE EMA TRANSPORTER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: 299121; CAB15405.1; -
 DR Subtilist; BG14088; YVBY.
 DR InterPro: IPR000620; DUF6.
 DR Pfam: PF00892; DUF6; 2.

KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
 FT TRANSMEM 11 31 POTENTIAL.
 FT TRANSMEM 37 57 POTENTIAL.
 FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 97 117 POTENTIAL.
 FT TRANSMEM 126 146 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 180 200 POTENTIAL.
 FT TRANSMEM 217 237 POTENTIAL.
 FT TRANSMEM 244 264 POTENTIAL.
 FT TRANSMEM 265 285 POTENTIAL.
 SQ SEQUENCE 305 AA; 32820 MW; 203B2E4E7138670B CRC64;

Query Match 50.0%; Score 43.5; DB 1; Length 305;
 Best Local Similarity 46.7%; Pred. No. 28;
 Matches 7; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

OY 2 SMLVFL-LLMGVTV 15
 Db 10 ALLAFVIMGVNM 24

RESULT 7
 YORS_EAV

ID YORS_EAV STANDARD; PRT; 255 AA.

AC P28995;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE HYPOTHETICAL 28.7 KDA PROTEIN (ORF5).
 OS Equine arteritis virus (EAV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 CC Arteriviridae; Arterivirus.
 OX NCBI_TaxID:11047;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BUCYRUS;
 RX MEDLINE:91237805; PubMed:1851863;
 RA den Boon J.A., Snijder E.J., Chirnside E.D., de Vries A.A.F.,
 RA Horzinek M.C., Spaan W.J.M.;
 RT "Equine arteritis virus is not a togavirus but belongs to the
 RT coronaviruslike superfamily".
 RL J. Virol. 65:2910-2920(1991).
 CC -1- FUNCTION: POSSIBLE ENVELOPE PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X53459; CA37544.1; -
 DR PIR: F39925; F39925.
 DR InterPro: IPR001332; Arteri_glycop.
 DR InterPro: IPR003241; Equine_virus_ORF5.
 DR Pfam: PF00951; Arteri_glycop; 1 ORF5.
 DR Prodom: PD002371; Equine_virus_ORF5; 1.
 KW Hypothetical protein; Envelope protein.
 FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 255 AA; 28672 MW; 4310EBED1134B399 CRC64;

Query Match 49.4%; Score 43; DB 1; Length 255;
 Best Local Similarity 58.3%; Pred. No. 28;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 SMLVFL-LLMG 12
 Db 2 LSMIVLFLMG 13

RESULT 8
 ADT_CHLRE STANDARD; PRT; 308 AA.
 AC P27080;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE ADP CARRIER PROTEIN (ADP/ATP TRANSLOCASE) (ADENINE NUCLEOTIDE
 DE TRANSLOCATOR) (ANT).
 GN APT.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 CC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID:3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FUD4-R2;
 RX MEDLINE:9320487; PubMed:8455552;
 RA Sharpe J.A., Day A.;
 RT "Structure, evolution and expression of the mitochondrial ADP/ATP
 RT translocator gene from Chlamydomonas reinhardtii".
 RL Mol. Gen. Genet. 237:134-144(1993).
 CC -1- FUNCTION: CATALYZES THE EXCHANGE OF ADP AND ATP ACROSS THE
 CC MITOCHONDRIAL INNER MEMBRANE.

```
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X65194; CAA46311.1; -.
CC PIR: S30259; S30259.
CC InterPro: IPR002067; Mit_carrier.
CC InterPro: IPR001993; Mitoch_carrier.
CC Pfam: PF00153; mito_carr; 3.
CC PRINTS: PR00926; MITOCARRIER.
CC PRINTS: PR00927; ADPTRNSLCASE.
CC PROSITE: PS00215; MITOCH_CARRIER_2.
CC KW Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport.
FT TRANSSEM 12 29 1 (POTENTIAL).
FT TRANSSEM 74 92 2 (POTENTIAL).
FT TRANSSEM 116 133 3 (POTENTIAL).
FT TRANSSEM 178 197 4 (POTENTIAL).
FT TRANSSEM 217 234 5 (POTENTIAL).
FT TRANSSEM 273 291 6 (POTENTIAL).
SQ SEQUENCE 308 AA; 33528 MW; D477CF0E72B7A53F CRC64;

Query Match 49.4%; Score 43; DB 1; Length 308;
Best Local Similarity 53.3%; Pred. No. 32;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 SMLVFLILMGVTWG 16
Db 212 NFLAFLLGNGTTIG 226

RESULT 9
ID OPSD TODPA STANDARD; PRT; 447 AA.
AC P31356;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN.
GN RHO.
OS Rhodops pacificus (Japanese flying squid).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Teuthoidea;
OC Oegopsida; Ommastrephidae; Todarodes.
OX NCBI_TaxID=6637;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-16; 128-139 AND 302-312.
RC TISSUE-Retina;
RA MEDLINE=93154520; PubMed=6428633;
RA Hara-Nishimura I., Kondo M., Nishimura M., Hara R., Hara T.;
RT "Cloning and nucleotide sequence of cDNA for rhodopsin of the squid
RT Todarodes pacificus.";
RL FEBS Lett. 317:5-11(1993).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL. THIS RECEPTOR IS COUPLED TO THE ACTIVATION
CC OF PHOSPHOLIPASE C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PMW: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X70498; CAA49906.1; -.
CC PIR: S29483; S29483.
CC GCRDB: GCR_0576; -.
CC InterPro: IPR000276; GPCR_Rhodopsn.
CC InterPro: IPR000216; Molisc_rhodopsn_C.
CC InterPro: IPR001760; Opsin.
CC Pfam: PF00001; 7tm.1; 1.
CC PRINTS: PR02162; Rhodopsin_C; 1.
CC PRINTS: PR00237; GPCR_RHODOPSIN.
CC PRINTS: PR00238; OPSIN.
CC PRINTS: PR00239; RHODOPSNFALL.
CC PROSITE: PS00237; G-PROTEIN_RECEP_FL_1; 1.
CC PROSITE: PS0262; G-PROTEIN_RECEP_FL_2; 1.
CC PROSITE: PS00238; OPSIN; 1.
CC KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
CC Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
FT INIT_MET 0 0
FT DOMAIN 1 32 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 33 57 1 (POTENTIAL).
FT DOMAIN 58 69 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 70 96 2 (POTENTIAL).
FT DOMAIN 97 110 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 111 130 3 (POTENTIAL).
FT DOMAIN 131 150 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 151 174 4 (POTENTIAL).
FT DOMAIN 175 198 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 199 226 5 (POTENTIAL).
FT DOMAIN 227 260 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 261 284 6 (POTENTIAL).
FT DOMAIN 285 292 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 293 317 7 (POTENTIAL).
FT DOMAIN 318 447 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 7 7 N-LINKED (GLUCNA...).
FT DISULFID 107 185 BY SIMILARITY.
FT BINDING 304 304 RETINAL CHROMOPHORE (BY SIMILARITY).
FT LIPID 335 335 PALMITATE (BY SIMILARITY).
FT LIPID 336 336 PALMITATE (BY SIMILARITY).
FT DOMAIN 370 380 MET-RICH.
FT DOMAIN 381 447 GLN/PRO-RICH.
SQ SEQUENCE 447 AA; 49704 MW; ACTACD7FB15A01C1 CRC64;

Query Match 49.4%; Score 43; DB 1; Length 447;
Best Local Similarity 38.5%; Pred. No. 43;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 MLVFLILMGVTW 15
Db 153 IMIIFWLMSTVLM 165

RESULT 10
ID PACO_SALTY STANDARD; PRT; 304 AA.
AC O30646;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PACO PROTEIN.
GN PACO.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneilla.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-ATCC 14028S;
RX MEDLINE:98380520; PubMed:9712687;
RA Gunn J.S., Belden W.J., Miller S.L.;
RT "Identification of Phor-PhoQ activated genes within a duplicated
RL region of the Salmonella typhimurium chromosome."
CC Microb. Pathog. 25:77-90(1998).
CC -1- SIMILARITY: BELONGS TO THE EMA TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF013775; AAB82452.1; .
DR StryGene: SG10678; PAGO.
DR InterPro: IPR000620; DUF6.
DR Pfam: PF00892; DUF6. 2.
KM Transmembrane.
FT TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 95 115 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 150 170 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 246 266 POTENTIAL.
FT TRANSMEM 267 287 POTENTIAL.
SQ SEQUENCE 304 AA; 33622 MW; B368153FD6EBFB08 CRC64;

Query Match 48.9%; Score 42.5; DB 1; Length 304;
Best Local Similarity 38.9%; Pred. No. 37;
Matches 7; Conservative 6; Mismatches 2; Indels 3; Gaps 1;

Oy 1 MSMLVFFLL---KGVTV 15
Db 4 VSIILEFVLVSLTWGTW 21

RESULT 11
SECU_METVA
ID SECU_METVA STANDARD: PRT; 438 AA.
AC P28541.
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PREPROTEIN TRANSLOCASE SECU SUBUNIT.
GN SECU.
OS Methanococcus vannielii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID:2187;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:92110434; PubMed:1764515;
RA Auer J., Spicker G., Boeck A.;
RT "Presence of a gene in the archaeobacterium Methanococcus vannielii
RL homologous to secY of Escherichia coli."
RT Biochimie 73:683-688(1991).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. INTERACTS WITH SECA AND SECE
CC TO ALLOW THE TRANSLOCATION OF PROTEINS ACROSS THE PLASMA MEMBRANE,
CC BY FORMING PART OF A CHANNEL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE SECU/SEC1-ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X62045; CAA43978.1; .
DR PIR: S24065; S24065.
DR InterPro: IPR002308; SecY.
DR Pfam: PF00344; secy. 1.
DR PROSITE: PS00755; SECY_1;
DR PROSITE: PS00756; SECY_2; 1.
KM Protein transport; Translocation; Transmembrane.
FT TRANSMEM 28 48 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 136 156 POTENTIAL.
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 205 225 POTENTIAL.
FT TRANSMEM 257 277 POTENTIAL.
FT TRANSMEM 291 311 POTENTIAL.
FT TRANSMEM 316 336 POTENTIAL.
FT TRANSMEM 380 400 POTENTIAL.
SQ SEQUENCE 438 AA; 47615 MW; A89B258DB44E946A CRC64;

Query Match 48.3%; Score 42; DB 1; Length 438;
Best Local Similarity 46.2%; Pred. No. 58;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 3 MLVFFLLKGVTV 15
Db 323 MNVIFCLFGLFW 335

RESULT 12
OPSD_OCTDO
ID OPSD_OCTDO STANDARD: PRT; 455 AA.
AC P09241.
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN.
GN RHO.
OS Octopus dofleini (Giant octopus).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Octopoda;
OC Euprymata; Octopodidae; Octopus.
OX NCBI_TaxID:6644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:88211878; PubMed:3366250;
RA Ovchinnikov Y.A., Abdulaev N.G., Zolotarev A.S., Artamonov I.D.,
RA Bessalov I.A., Dergachev A.E., Tsuda M.;
RT "Octopus rhodopsin. Amino acid sequence deduced from cDNA."
RT FEBS Lett. 232:69-72(1988).
RN [2]
RP REVISION TO 399.
RA Abdulaev N.G.;
RL Submitted (OCT-1988) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.
CC -1- PTR: SOME OR ALL OF THE CARBOXYL-TERMINAL, SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```


CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X07797; CAA30644.1; -.
 DR PIR; S00610; OOCOG.
 DR GCRDB; GCR 0100; -.
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR InterPro: IPR000216; Molisc_rhodopsn_C.
 DR InterPro: IPR001760; Opsin.
 DR Pfam; PF00001; 7tm_1; 1.
 DR Pfam; PF02162; Rhodopsin_C; 1.
 DR PRINTS; PR00237; GPCR_RHODOPSIN.
 DR PRINTS; PR00238; OPSIN.
 DR PRINTS; PR00239; RHODOPSNTAIL.
 DR PROSITE; PS00237; G-PROTEIN_RECEP_FL_1; 1.
 DR PROSITE; PS50262; G-PROTEIN_RECEP_FL_2; 1.
 DR PROSITE; PS00238; OPSIN; 1.
 KM Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 KM Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor;
 KM Repeat.
 FT DOMAIN 1 34 EXTRACELLULAR.
 FT TRANSMEM 35 59 1 (POTENTIAL).
 FT DOMAIN 60 71 CYTOPLASMIC.
 FT TRANSMEM 72 98 2 (POTENTIAL).
 FT DOMAIN 99 112 EXTRACELLULAR.
 FT TRANSMEM 113 132 3 (POTENTIAL).
 FT DOMAIN 133 152 CYTOPLASMIC.
 FT TRANSMEM 153 176 4 (POTENTIAL).
 FT DOMAIN 177 200 EXTRACELLULAR.
 FT TRANSMEM 201 228 5 (POTENTIAL).
 FT DOMAIN 229 262 CYTOPLASMIC.
 FT TRANSMEM 263 286 6 (POTENTIAL).
 FT DOMAIN 287 294 EXTRACELLULAR.
 FT TRANSMEM 295 319 7 (POTENTIAL).
 FT DOMAIN 320 455 CYTOPLASMIC.
 FT CARBOHYD 9 9 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT DISULFID 109 187 BY SIMILARITY.
 FT BINDING 306 306 RETINAL CHROMOPHORE.
 FT LIPID 337 337 PALMITATE (BY SIMILARITY).
 FT LIPID 338 338 PALMITATE (BY SIMILARITY).
 FT DOMAIN 370 380 MET-RICH.
 FT DOMAIN 381 455 GLN/PRO-RICH.
 FT TRANSMEM 395 431 6 X 5 AA REPEATS OF G-Y-P-P-Q.
 FT REPEAT 395 399 1-1.
 FT REPEAT 400 404 1-2.
 FT REPEAT 412 416 2-1.
 FT REPEAT 417 421 2-2.
 FT REPEAT 422 426 3-1.
 FT REPEAT 427 431 3-2.
 SQ SEQUENCE 455 AA; 50488 MW; AECFAFC30ED1346C CRC64;
 Query Match 48.3%; Score 42; DB 1; Length 455;
 Best Local Similarity 23.1%; Pred. No. 60;
 Matches 3; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
 OY 3 MLVFLILMGVTW 15
 DB 155 LMIIFVMMSTIV 167
 RESULT 13
 XASA_ECO57 STANDARD; PRT; 511 AA.
 AC P58229;
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE AMINO ACID ANTIporter (EXTREME ACID SENSITIVITY PROTEIN).
 GN XASA OR ACSA OR GADC OR 22216 OR ECS2097.
 OS Escherichia coli O157:H7.

CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC NCBI_TaxID=83334;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouls K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT *Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.*
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT *Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.*;
 RL DNA Res. 8:11-22(2001).
 CC -1- FUNCTION: PROBABLE AMINO ACID TRANSPORTER. INVOLVED IN EXTREME
 CC ACID RESISTANCE (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE XASA FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE005356; AAC56276.1; -.
 DR EMBL; AP002557; BAB35520.1; -.
 KM Transmembrane; Inner membrane; Transport; Amino-acid transport.
 FT TRANSMEM 15 35 POTENTIAL.
 FT TRANSMEM 42 62 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 128 148 POTENTIAL.
 FT TRANSMEM 158 178 POTENTIAL.
 FT TRANSMEM 201 221 POTENTIAL.
 FT TRANSMEM 240 260 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.
 FT TRANSMEM 336 356 POTENTIAL.
 FT TRANSMEM 367 387 POTENTIAL.
 FT TRANSMEM 413 433 POTENTIAL.
 FT TRANSMEM 446 466 POTENTIAL.
 SQ SEQUENCE 511 AA; 55102 MW; 85DF72BD529CE33 CRC64;
 Query Match 48.3%; Score 42; DB 1; Length 511;
 Best Local Similarity 66.7%; Pred. No. 65;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 OY 4 LVVFLILMGVTW 15
 DB 42 LVVFLILMGVTW 53
 RESULT 14
 XASA_ECOLI STANDARD; PRT; 511 AA.
 ID XASA_ECOLI
 AC P39183; P76131; P77384;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, last annotation update)
 DE AMINO ACID ANTIPORTER (EXTREME ACID SENSITIVITY PROTEIN).
 GN XASA OR ACBA OR GADC OR B1492.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID:562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655.
 RX MEDLINE:97426617; PubMed:9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE:97251357; PubMed:9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
 RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubramanian S.,
 RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
 RA Yamamoto Y., Horieuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map";
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 402-490 FROM N.A.
 RC STRAIN-K12 / MC4100;
 RA Hersch B.M., Blankenhorn D., Slonczewski J.L.;
 RT Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROBABLE AMINO ACID TRANSPORTER. INVOLVED IN EXTREME
 CC ACID RESISTANCE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE XASA FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000246; AAC74565.1; -;
 DR EMBL: D90790; BAA15156.1; ALT_SEQ.
 DR EMBL: D90791; BAA15162.1; ALT_SEQ.
 DR EMBL: U13204; AAB17694.1; -;
 DR EcoGene: EG12436; xasa.
 DR InterPro: IPR002293; AA_rel_permease-1.
 DR InterPro: IPR002027; Amino_acid_permease.
 DR Pfam: PF00324; aa-permeases; 1.
 KW Transmembrane; Inner membrane; Transport; Amino-acid transport;
 KW Complete proteome.
 FT TRANSMEM 15 35
 FT TRANSMEM 42 62 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 128 148 POTENTIAL.
 FT TRANSMEM 158 178 POTENTIAL.
 FT TRANSMEM 201 221 POTENTIAL.
 FT TRANSMEM 240 260 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.
 FT TRANSMEM 336 356 POTENTIAL.
 FT TRANSMEM 367 387 POTENTIAL.
 FT TRANSMEM 413 433 POTENTIAL.
 FT TRANSMEM 446 466 POTENTIAL.
 FT TRANSMEM 511 AA: 55076 MW: 85DF72B82816CE33 CRC64;
 SQ SEQUENCE

Query Match 48.3%; Score 42; DB 1; Length 511;
 Best Local Similarity 66.7%; Pred. No. 65;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Oy 4 LVFFLLMGVTV 15
 Db 42 LVFFLLGCIIM 53
 RESULT 15
 XASA_SHIFL STANDARD; PRT; 511 AA.
 ID XASA_SHIFL
 AC 054152;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE AMINO ACID ANTIPORTER (EXTREME ACID SENSITIVITY PROTEIN).
 GN XASA OR GADC.
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Shigella.
 OX NCBI_TaxID:623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-M25-8A;
 RX MEDLINE:97039712; PubMed:8885264;
 RA Waterman S.R., Small P.L.C.;
 RT "Identification of sigma S-dependent genes associated with the
 RT stationary-phase acid-resistance phenotype of Shigella flexneri.";
 RL Mol. Microbiol. 21:925-940(1996).
 CC -1- FUNCTION: PROBABLE AMINO ACID TRANSPORTER. INVOLVED IN EXTREME
 CC ACID RESISTANCE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE XASA FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U46133; AAD14843.1; -;
 DR InterPro: IPR002293; AA_rel_permease-1.
 DR InterPro: IPR002027; Amino_acid_permease.
 DR Pfam: PF00324; aa-permeases; 1.
 KW Transmembrane; Inner membrane; Transport; Amino-acid transport;
 FT TRANSMEM 15 35
 FT TRANSMEM 42 62 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 128 148 POTENTIAL.
 FT TRANSMEM 158 178 POTENTIAL.
 FT TRANSMEM 201 221 POTENTIAL.
 FT TRANSMEM 240 260 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.
 FT TRANSMEM 336 356 POTENTIAL.
 FT TRANSMEM 367 387 POTENTIAL.
 FT TRANSMEM 413 433 POTENTIAL.
 FT TRANSMEM 446 466 POTENTIAL.
 FT TRANSMEM 511 AA: 55020 MW: 8369AE15AF1523D CRC64;
 SQ SEQUENCE
 Query Match 48.3%; Score 42; DB 1; Length 511;
 Best Local Similarity 66.7%; Pred. No. 65;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Oy 4 LVFFLLMGVTV 15
 Db 42 LVFFLLGCIIM 53

```
RESULT 16
DAL5_YEAST STANDARD: PRT: 543 AA.
ID DAL5_YEAST
AC P15365;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ALLANTOATE PERMEASE
GN DAL5 OR UREPI OR YUR152W OR J2230.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88086880; PubMed=3275614;
RA Rai R., Genbauffe F.S., Cooper T.G.;
RT "Structure and transcription of the allantate permease gene (DAL5)
RT from Saccharomyces cerevisiae.";
RL J. Bacteriol. 170:266-271(1986).
RN [2]
RP SEQUENCE FROM N.A.
RA Scarcez T.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE ALLANTOATE TRANSPORT SYSTEM.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE ALLANTOATE PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24098; AAA34555.1; -.
DR EMBL; Z49652; CA89685.1; -.
DR PIR; A28671; A28671.
DR SGD; S0003913; DAL5.
KW Transport; Transmembrane.
FT TRANSMEM 81 97 POTENTIAL.
FT TRANSMEM 124 145 POTENTIAL.
FT TRANSMEM 155 171 POTENTIAL.
FT TRANSMEM 179 200 POTENTIAL.
FT TRANSMEM 214 237 POTENTIAL.
FT TRANSMEM 249 269 POTENTIAL.
FT TRANSMEM 318 342 POTENTIAL.
FT TRANSMEM 353 377 POTENTIAL.
FT TRANSMEM 390 411 POTENTIAL.
FT TRANSMEM 418 435 POTENTIAL.
FT TRANSMEM 454 472 POTENTIAL.
FT TRANSMEM 483 504 POTENTIAL.
FT CARBOHYD 10 10 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 19 19 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 543 AA; 60850 MW; 9B70CAF3FAD0BB8 CRC64;

Query Match 48.3%; Score 42; DB 1; Length 543;
Best Local Similarity 54.5%; Pred. No. 68;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
ID NTDO_CAEEL STANDARD: PRT: 615 AA.
AC 003614; O9YTKO;
DT 01-FEB-1994 (Rel. 28, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SODIUM-DEPENDENT DOPAMINE TRANSPORTER (DA TRANSPORTER) (DAT).
GN T23G5.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98440631; PubMed=9765501;
RA Jayanthi L.D., Apparsundaram S., Malone M.D., Ward E., Miller D.M.,
RA Eppler M., Blakely R.D.;
RT "The Caenorhabditis elegans gene T23G5.5 encodes an antidepressant-and
RT cocaine-sensitive dopamine transporter.";
RL Mol. Pharmacol. 54:601-609(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=BRISTOL NZ;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showken R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP REVIEWS.
RA Dublin R.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: TERMINATES THE ACTION OF DOPAMINE BY ITS HIGH
CC AFFINITY SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- MISCELLANEOUS: THIS PROTEIN IS THE TARGET OF PSYCHOMOTOR
CC STIMULANTS SUCH AS AMPHETAMINES OR COCAINE.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF079899; AAC83661.1; -.
DR EMBL; Z19158; CAA79575.2; -.
DR EMBL; Z19156; CAA79575.2; JOINED.
DR EMBL; Z19156; CAA79564.2; -.
DR EMBL; Z19158; CAA79564.2; JOINED.
DR PIR; S28306; S28306.
DR WormPep; T23G5.5; CE25124.
DR InterPro; IPR000175; Na_neurotran_symport.
DR Pfam; PF00209; SNF; 1.
DR PRINTS; PR00176; NANUSMPORT.
DR PROSITE; PS00610; NA_NEUROTRAN_SYM_1; 1.
DR PROSITE; PS00754; NA_NEUROTRAN_SYM_2; 1.
DR PROSITE; PS0267; NA_NEUROTRAN_SYM_3; 1.
KW Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
KW Symport.
FT DOMAIN 1 48 CYTOPLASMIC (POTENTIAL).
```

```
FT TRANSMEM 49 69 1 (POTENTIAL).
FT TRANSMEM 74 94 2 (POTENTIAL).
FT TRANSMEM 125 145 3 (POTENTIAL).
FT DOMAIN 146 232 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 233 253 4 (POTENTIAL).
FT TRANSMEM 258 278 5 (POTENTIAL).
FT TRANSMEM 344 364 6 (POTENTIAL).
FT TRANSMEM 371 391 7 (POTENTIAL).
FT TRANSMEM 393 413 8 (POTENTIAL).
FT TRANSMEM 440 460 9 (POTENTIAL).
FT TRANSMEM 462 482 10 (POTENTIAL).
FT TRANSMEM 515 535 11 (POTENTIAL).
FT TRANSMEM 550 570 12 (POTENTIAL).
FT DOMAIN 571 615 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 615 AA: 69265 MM: 0DA2876B3AC8049 CXC64;

Query Match 48.3%; Score 42; DB 1; Length 615;
Best Local Similarity 35.7%; Pred. No. 75;
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

OY 3 MSLVFLILMGVTWG 16
Db 463 LIMELIITGTTWG 476

RESULT 18
MXIJ_SHIFL STANDARD; PRT; 241 AA.
ID MXIJ_SHIFL
AC 006081;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE LIPOPROTEIN MXIJ PRECURSOR.
GN MXIJ.
OS Shigella flexneri.
OC Plasmid 210 kb invasion pMR100.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Shigella.
OX NCBI_Taxid=623;
RN 11
RP SEQUENCE FROM N.A.
RC MEDLINE:93077448; PubMed:1332940;
RX Allabout A., Sansonetti P.J., Parsot C.;
RT "MXIJ, a lipoprotein involved in secretion of Shigella ipa invasins,
RT is homologous to YscJ, a secretion factor of the Yersinia Yop
RT proteins";
RL J. Bacteriol. 174:7661-7669(1992).
CC -1- FUNCTION: INVOLVED IN THE SECRETION OF THE IPA ANTIGENS.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE YSCJ FAMILY OF LIPOPROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M98390; AAA26532.1; -.
CC PIR: C45271; C45271.
CC InterPro: IPR003282; SecYIIPMK.
CC InterPro: IPR002920; YscJ_F11F.
CC Pfam: PF01514; YscJ_F11F; 1.
CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
CC Virulence: Plasmid; Transport; Protein transport; Outer membrane;
CC Signal; Lipoprotein.
FT SIGNAL 1 17 POTENTIAL.
```

```
FT CHAIN 18 241 LIPOPROTEIN MXIJ.
FT LIPID 18 18 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 241 AA: 27509 MM: 3D20D62033A30898 CXC64;

Query Match 47.1%; Score 41; DB 1; Length 241;
Best Local Similarity 50.0%; Pred. No. 49;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MSLVFLILMGVTWG 16
Db 219 MAVLVVILLVWAFKFG 234

RESULT 19
MXIJ_SHISO STANDARD; PRT; 241 AA.
ID MXIJ_SHISO
AC 055288;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE LIPOPROTEIN MXIJ PRECURSOR.
GN MXIJ.
OS Shigella sonnei.
OC Plasmid.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Shigella.
OX NCBI_Taxid=624;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-HW383;
RA Arakawa E., Kato J.I., Ito K.I., Watanabe H.;
RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE SECRETION OF THE IPA ANTIGENS.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE YSCJ FAMILY OF LIPOPROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: D50601; BAA09148.1; -.
CC InterPro: IPR003282; SecYIIPMK.
CC InterPro: IPR002920; YscJ_F11F.
CC Pfam: PF01514; YscJ_F11F; 1.
CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
CC Virulence: Plasmid; Transport; Protein transport; Outer membrane;
CC Signal; Lipoprotein.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 241 LIPOPROTEIN MXIJ.
FT LIPID 18 18 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 241 AA: 27523 MM: 1FB3D62033A31B20 CXC64;

Query Match 47.1%; Score 41; DB 1; Length 241;
Best Local Similarity 50.0%; Pred. No. 49;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MSLVFLILMGVTWG 16
Db 219 MAVLVVILLVWAFKFG 234

RESULT 20
YIJE_ECOLI STANDARD; PRT; 301 AA.
ID YIJE_ECOLI
AC P32667;
DT 01-OCT-1993 (Rel. 27, Created)
```

DT 15-DEC-1998 (Rel. 37, last sequence update)
DE 20-AUG-2001 (Rel. 40, last annotation update)
DE HYPOPHYSICAL PROTEIN YIUE.
GN YIUE OR B3943.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=94089392; PubMed=8265357;
RA Blatner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes."
RL Nucleic Acids Res. 21:5408-5417(1993).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE EMA TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U00006; AAC3049.1; ALT_INIT.
DR EMBL: AE000468; AAC76925.1; ALT_INIT.
DR EcoGene: EG11902; Y13E.
DR InterPro: IPR000620; DUF6.
DR Pfam: PF00892; DUF6; 2.
DR
KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 219 239 POTENTIAL.
FT TRANSMEM 246 266 POTENTIAL.
FT TRANSMEM 269 289 POTENTIAL.
SQ SEQUENCE 301 AA; 32865 MM; 1A822058FDA28696 CRC64;

Query Match 47.1%; Score 41; DB 1; Length 301;
Best Local Similarity 46.7%; Pred. No. 59;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MSMLVFLIMGVTW 15
Db 12 ISGLVLTLMYSW 26

RESULT 21
ID OPSD_LIZSA STANDARD; PRT; 353 AA.
AC O9YG27;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, last sequence update)
DT 20-AUG-2001 (Rel. 40, last annotation update)
DE RHODOPSIN.
GN RHO.
OS Liza saliens (leaping mullet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Mugilomorpha; Mugilidae;
OC Liza.
OX NCBI_TaxID=48192;
RN
RP SEQUENCE FROM N.A.

RC TISSUE=Retina;
RA Archer S.N., Hirano J.;
RT "Comparative analysis of opsins in Mediterranean coastal fish."
RL Submitted (JAN-1999) to the EMBL/Genbank/DBD databases.
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y18670; CAA77252.1;
DR InterPro: IPR000276; GPCR_Rhodopsin.
DR InterPro: IPR001760; Opsin.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCR_RHODOPSIN.
DR PRINTS: PR00238; RHODOPSIN.
DR PRINTS: PR00579; RHODOPSIN.
DR PROSITE: PS00237; G-PROTEIN_RECP_F1_1; 1.
DR PROSITE: PS50262; G-PROTEIN_RECP_F1_2; 1.
DR PROSITE: PS00238; OPSIN; 1.
DR
KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
FT TRANSMEM 1 36 POTENTIAL.
FT TRANSMEM 37 61 POTENTIAL.
FT TRANSMEM 62 73 POTENTIAL.
FT TRANSMEM 74 98 POTENTIAL.
FT TRANSMEM 99 113 POTENTIAL.
FT TRANSMEM 114 133 POTENTIAL.
FT TRANSMEM 134 152 POTENTIAL.
FT TRANSMEM 153 176 POTENTIAL.
FT TRANSMEM 177 202 POTENTIAL.
FT TRANSMEM 203 230 POTENTIAL.
FT TRANSMEM 231 252 POTENTIAL.
FT TRANSMEM 253 276 POTENTIAL.
FT TRANSMEM 277 284 POTENTIAL.
FT TRANSMEM 285 309 POTENTIAL.
FT TRANSMEM 310 353 POTENTIAL.
FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT BINDING 296 296 RETINAL CHROMOPHORE (BY SIMILARITY).
FT LIPID 322 322 PALMITATE (BY SIMILARITY).
FT DISULFID 110 187 BY SIMILARITY.
SQ SEQUENCE 353 AA; 39566 MM; A155664649079E3C CRC64;

Query Match 47.1%; Score 41; DB 1; Length 353;
Best Local Similarity 47.1%; Pred. No. 66;
Matches 8; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

OY 3 MLVVFLLIM---GVTW 15
Db 257 MVVAFLLICWCPYAGVAVW 273

RESULT 22
ID OPSD_MUGCE STANDARD; PRT; 353 AA.
AC O9YG29;
DT 20-AUG-2001 (Rel. 40, Created)

DR InterPro: IPR003916; NADH:ub:oxidctse5.
 DR InterPro: IPR001750; Oxidored_q1.
 DR InterPro: IPR001516; Oxidored_q1_N.
 DR Pfam: PF00361; oxidored_q1; 1.
 DR Pfam: PF00662; oxidored_q1_N; 1.
 KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 FT CONFLICT 1 M -> A (IN REF. 1 AND 5).
 FT CONFLICT 206 206 G -> A (IN REF. 1 AND 5).
 FT CONFLICT 292 292 I -> S (IN REF. 1 AND 5).
 SO SEQUENCE 546 AA; 59047 MW; B7AA04BBB53B17DD CRC64;

Query Match 47.1%; Score 41; DB 1; Length 546;
 Best Local Similarity 75.0%; Pred. No. 93;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 8 LLMGVTV 15
 111111
 DB 179 LLMGVTV 186

RESULT 24
 ATIS_HUMAN STANDARD: PRT: 797 AA.
 ID ATIS_HUMAN
 AC P8196;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE IS (EC 3.6.3.13)
 DE (FRAGMENT).
 GN ATP11A OR ATP1S OR KIA1021.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE=9397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:197-205(1999).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O -> ADP + PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
 CC ATPASES). SUBFAMILY IV.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AB028944; BAA82973.1; -;
 DR MIM: 605868;
 DR InterPro: IPR001757; E1-E2_ATPase.
 DR InterPro: IPR001454; Hydrolase.
 DR Pfam: PF00702; Hydrolase; 1.
 DR PROSITE: PS00154; ATPase_E1_E2; 1.
 KM Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
 KW Multigene family.
 FT DOMAIN 1 1
 FT NON_TER 1 1
 FT TRANSSEM 13 12 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 36 35 POTENTIAL.
 FT TRANSSEM 544 544 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 545 545 POTENTIAL.
 FT DOMAIN 566 577 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 578 597 POTENTIAL.

FT DOMAIN 598 627 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 628 649 POTENTIAL.
 FT DOMAIN 650 663 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 664 686 POTENTIAL.
 FT DOMAIN 687 692 POTENTIAL.
 FT TRANSSEM 693 713 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 714 731 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 732 756 POTENTIAL.
 FT DOMAIN 757 797 CYTOPLASMIC (POTENTIAL).
 FT MOD_RES 77 77 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 488 488 MAGNESIUM (BY SIMILARITY).
 FT METAL 492 492 MAGNESIUM (BY SIMILARITY).
 SO SEQUENCE 797 AA; 91040 MW; E64E1AEC5CA31045 CRC64;

Query Match 46.6%; Score 40.5; DB 1; Length 797;
 Best Local Similarity 75.0%; Pred. No. 1,5e+02;
 Matches 9; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 5 VVFLTW-GVTW 15
 111111111
 DB 706 VVFLTW-GVTW 717

RESULT 25
 LSHB_TRIUV STANDARD: PRT: 141 AA.
 ID LSHB_TRIUV
 AC O46482;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE LUTROPIN BETA CHAIN PRECURSOR (LUTEINIZING HORMONE BETA SUBUNIT) (LSH-
 DE BETA) (LSH-B) (LH-B).
 GN LHB.
 OS Trichosurus vulpecula (Brush-tailed possum).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Metatheria; Dipodomorpha; Phalangeridae; Trichosurus.
 OX NCBI_TaxID=9337;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Pituitary;
 RX MEDLINE=96345424; PubMed=9680384;
 RA Harrison G.A., Deane E.M., Cooper D.W.;
 RT "cDNA cloning of luteinizing hormone subunits from brushtail possum
 RT and red kangaroo.";
 RL Mamm. Genome 9:638-642(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Lawrence S.B., McNatty K.P., Fidler A.E.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROMOTES SPERMATOGENESIS AND OVULATION BY STIMULATING
 CC THE TESTES AND OVARIES TO SYNTHESIZE STEROIDS.
 CC -1- SUBUNIT: HETERODIMER OF A COMMON ALPHA CHAIN AND A UNIQUE BETA
 CC CHAIN WHICH CONFERES BIOLOGICAL SPECIFICITY TO THYROTROPIN,
 CC LUTROPIN, FOLLITROPIN AND GONADOTROPIN.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOPROTEIN HORMONES BETA CHAIN
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF017448; AAC96019.1; -;
 DR EMBL: AF090388; AAC63526.1; -;
 DR HSSP: P01233; IHRP.
 DR InterPro: IPR000359; Gys_knot.
 DR InterPro: IPR001545; Glyco_hormone_beta.
 DR Pfam: PF00007; Gys_knot; 1.
 DR SMART: SMO0068; GHB; 1.

DR PROSITE: PS00261; GLYCO_HORMONE_BETA_1; 1.
DR PROSITE: PS00689; GLYCO_HORMONE_BETA_2; 1.
KW Hormone; Signal: Glycoprotein.
FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 141 LUTROPIN BETA CHAIN.
FT DISULFID 30 78 BY SIMILARITY.
FT DISULFID 44 93 BY SIMILARITY.
FT DISULFID 47 131 BY SIMILARITY.
FT DISULFID 55 109 BY SIMILARITY.
FT DISULFID 59 111 BY SIMILARITY.
FT DISULFID 114 121 BY SIMILARITY.
FT CARBOHYD 34 34 N-LINKED (GLCNAC, . . .) (POTENTIAL).
SQ SEQUENCE 141 AA; 15060 MW; CCGF98036B3C4EE0 CRC64;

Query Match 46.0%; Score 40; DB 1; Length 141;
Best Local Similarity 43.8%; Pred. No. 44;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSMLVFLLMGVTWG 16
Db 7 LTVLLLLLEGGSGWG 22

RESULT 26
YFHB_ECOLI
ID YFHB_ECOLI STANDARD; PRT: 211 AA.
AC P30133;
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 24.4 KDA PROTEIN IN PURI-DPJ INTERGENIC REGION.
GN YFHB OR B2560.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_TaxID:562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN:NM137;
RX MEDLINE:92292954; PubMed:1602968;
RA Poulsen L.K., Larsen N.W., Molin S., Andersson P.;
RT "Analysis of an Escherichia coli mutant strain resistant to the cell-
killing function encoded by the *gef* gene family.";
RL Mol. Microbiol. 6:895-905(1992).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN:K12;
RA Nashimoto H., Saito N.;
RT Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN:K12 / MG1655;
RX MEDLINE:97426617; PubMed:9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick N.A., Coeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X72336; CAAS1063.1; ALT_INT.
DR EMBL: D64044; BAA10910.1; ALT_INT.
DR EMBL: U36841; AAA79822.1; ALT_INT.
DR EMBL: AE000342; AAC75613.1; ALT_INT.

DR PIR: S20973; S20973.
DR Ecogene: EG11371; YFHB.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 211 AA; 24439 MW; 908E95FE54A3A2D CRC64;

Query Match 46.0%; Score 40; DB 1; Length 211;
Best Local Similarity 77.8%; Pred. No. 60;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 8 LLMGVTWG 16
Db 65 LLMGCTFG 73

RESULT 27
RFAL_KLEPN
ID RFAL_KLEPN STANDARD; PRT: 259 AA.
AC Q48475;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE O-ANTIGEN EXPORT SYSTEM PERMEASE PROTEIN RFBA.
GN RFBA.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Klebsiella.
OX NCBI_TaxID:573;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN:01:K20 / 889/50;
RX MEDLINE:95191401; PubMed:7533882;
RA Bronner D., Clarke B.R., Whitfield C.;
RT "Identification of an ATP-binding cassette transport system required
RT for translocation of 11polysaccharide O-antigen side-chains across
RT the cytoplasmic membrane of Klebsiella pneumoniae serotype O1.";
RL Mol. Microbiol. 14:505-519(1994).
CC -1- PUNCTON: MAY FORM AN ATP-DRIVEN O-ANTIGEN EXPORT APPARATUS, IN
CC ASSOCIATION WITH RFBA.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ABC-2 SUBFAMILY OF INTEGRAL MEMBRANE
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L31775; AAC98411.1; -
DR InterPro: IPR000412; ABC2_transport.
DR Pfam: PF01061; ABC2_membrane_1.
DR PROSITE: PS00890; ABC2_MEMBRANE_1.
KW Polysaccharide transport; Transport; Transmembrane; Inner membrane.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 111 131 POTENTIAL.
FT TRANSMEM 142 162 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 228 248 POTENTIAL.
SQ SEQUENCE 259 AA; 30036 MW; 074811E5BA5C493 CRC64;

Query Match 46.0%; Score 40; DB 1; Length 259;
Best Local Similarity 40.0%; Pred. No. 71;
Matches 8; Conservative 5; Mismatches 1; Indels 6; Gaps 1;

OY 3 MLVVFLLMGVT-----WG 16
Db 124 VIVVFLVYGMTPSLSWVG 143

RESULT 28
ID RCBL_CHRYI STANDARD: PRT: 278 AA.
AC P51762: 082945;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE REACTION CENTER PROTEIN L CHAIN (PHOTOSYNTHETIC REACTION CENTER L SUBUNIT).
GN POF1.
OS Chromatium vinosum.
OC Bacteria; Proteobacteria; gamma subdivision; Chromatiaceae;
OX NCBL_Taxid=1049;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-D:
RA Corson G.E., Nagashima K.V., Matsuura K., Sakuragi Y., Ruwanthi W., Qin H., Allen R., Knaf D.B.
RT "Primary structure of genes encoding light-harvesting and reaction center proteins from Chromatium vinosum."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 22-278 FROM N.A.
RA Nagashima K.V., Hiraishi A., Shimada K., Matsuura K.
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE REACTION CENTER IS A MEMBRANE-BOUND COMPLEX THAT MEDIATES THE INITIAL PHOTOCHEMICAL EVENT IN THE ELECTRON TRANSFER PROCESS OF PHOTOSYNTHESIS.
CC -1- SUBUNIT: REACTION CENTER IS COMPOSED OF FOUR BACTERIOCHLOROPHYLLS, TWO BACTERIOPHYLLOPHYSINS, TWO UBIQUINONES, ONE IRON, AND TWO HIGHLY HYDROPHOBIC POLYPEPTIDE CHAINS (DESIGNATED L AND M).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE REACTION CENTER L/M CHAINS / PSBA / PSBD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB011811; BAA32740.1; -
DR EMBL: D50647; BAA09314.1; -
DR HSSP: P02954; 4RCR.
DR InterPro: IPR000484; Photo_RC.
DR Pfam: PF00124; PhotoRC.1.
DR PRINTS: PR00256; REACTCENTRE.
DR PRODOM: PD000551; Photo_RC.1.
DR PROSITE: PS00244; REACTION_CENTER.1.
KW Transmembrane; Electron transport; Photosynthesis; Reaction center; Iron; Magnesium.
FT TRANSMEM 33 56 POTENTIAL.
FT TRANSMEM 90 118 POTENTIAL.
FT TRANSMEM 121 146 POTENTIAL.
FT TRANSMEM 176 205 POTENTIAL.
FT TRANSMEM 231 257 POTENTIAL.
FT METAL 179 179 MAGNESIUM (BACTERIOCHLOROPHYLL B) (BY SIMILARITY).
FT METAL 196 196 IRON (NON HAEM) (BY SIMILARITY).
FT METAL 236 236 IRON (NON HAEM) (BY SIMILARITY).
FT BINDING 222 222 QUINONE B (BY SIMILARITY).
FT CONFLICT 46 46 G -> V (IN REF. 2).
FT CONFLICT 130 130 A -> P (IN REF. 2).
FT CONFLICT 172 172 H -> N (IN REF. 2).
FT CONFLICT 200 200 I -> N (IN REF. 2).
FT CONFLICT 274 274 E -> N (IN REF. 2).
SQ SEQUENCE 278 AA; 31195 MW; 3C90852621A674E7 CRC64;

Query Match 46.0%; Score 40; DB 1; Length 278;
Best Local Similarity 46.7%; Pred. No. 75;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 2 SMLVFLLMGVTWG 16
Db 43 ALLGVLLVMGATIG 57
RESULT 29
ID OPSD_ICTPU STANDARD: PRT: 304 AA.
AC 042268;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHODOPSIN (FRAGMENT).
GN RHO.
OS Ictalurus punctatus (Channel catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Siluriformes; Ictaluridae; Ictalurus.
OX NCBL_Taxid=7998;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-97477428; PubMed-9334384;
RA Blackshaw S., Snyder S.H.;
RT "Paralopsin, a novel catfish opsin localized to the paraneal organ, defines a new gene family."
RL J. Neurosci. 17:8083-8092(1997).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES VISION IN DIM LIGHT.
CC -1- PPM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS. OPSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF028016; AAB84052.1; -
DR GCRDB: GCR_2524; -
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR InterPro: IPR001760; Opsin.
DR Pfam: PF00001; 7tm.1; 1.
DR PROSITE: PS00237; G_PROTEIN_RECPT_F1.1;
DR PROSITE: PS0262; G_PROTEIN_RECPT_F1.2; 1.
KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision; Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
FT NON_TER 1 1
FT DOMAIN 1 13 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 14 38 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 39 50 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 51 75 2 (POTENTIAL).
FT DOMAIN 76 90 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 91 110 3 (POTENTIAL).
FT DOMAIN 111 129 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 130 153 4 (POTENTIAL).
FT DOMAIN 154 179 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 180 207 5 (POTENTIAL).
FT DOMAIN 208 229 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 230 253 6 (POTENTIAL).

```

FT DOMAIN 254 261 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 262 286 7 (POTENTIAL).
FT DOMAIN 287 >304 CYTOPLASMIC (POTENTIAL).
FT DISULFID 87 164 BY SIMILARITY.
FT BINDING 273 273 RETINAL CHROMOPHORE (BY SIMILARITY).
FT LIPID 299 299 PALMITATE (BY SIMILARITY).
FT LIPID 300 300 PALMITATE (BY SIMILARITY).
FT NON_TER 304 304
SQ SEQUENCE 304 AA: 34406 MW: BIE9A76977136F46 CRC64:

Query Match 46.0% Score 40; DB 1; Length 304;
Best Local Similarity 41.2% Pred. No. 80;
Matches 7; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 3 MLVFFLLM---GVTW 15
Db 234 MVIAFLICWCPYAGVAV 250

RESULT 30
ADT_KLULA STANDARD: PRT: 305 AA.
ID ADT_KLULA
AC P49382;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE ADP-ATP CARRIER PROTEIN (ADP/ATP TRANSLOCASE) (ADENINE NUCLEOTIDE
DE TRANSLOCATOR) (ANT).
GN AAC.
OS Kluyveromyces fragilis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID:28985;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:95254664; PubMed:7736606;
RA Viola A.M., Galeotti C.L., Goffrini P., Ficarelli A., Ferrero I.;
RA "A Kluyveromyces fragilis gene homologue to AAC2 complements the
RT Saccharomycetes cerevisiae op1 mutation."
RL Curr. Genet. 27:229-233(1995).
CC -1- FUNCTION: CATALYZES THE EXCHANGE OF ADP AND ATP ACROSS THE
CC MITOCHONDRIAL INNER MEMBRANE.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L33797; AAC41655.1; -.
DR InterPro: IPR002067; MLC_carrier.
DR InterPro: IPR001993; Mitoch_carrier.
DR Pfam: PF00153; mito_carrier_3.
DR PRINTS: PR00926; MITOCARRIER.
DR PRINTS: PR00927; ADPTRNSLCASE.
DR PROSITE: PS00215; MITOCH_CARRIER_2.
KW Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport.
FT TRANSMEM 14 31 1 (POTENTIAL).
FT TRANSMEM 76 94 2 (POTENTIAL).
FT TRANSMEM 118 135 3 (POTENTIAL).
FT TRANSMEM 179 198 4 (POTENTIAL).
FT TRANSMEM 218 235 5 (POTENTIAL).
FT TRANSMEM 274 292 6 (POTENTIAL).
SQ SEQUENCE 305 AA: 33094 MW: 7FB95BCC18481A1 CRC64:

```

```

Query Match 46.0% Score 40; DB 1; Length 305;
Best Local Similarity 53.8% Pred. No. 81;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 SMLVFFLLMGVTWG 16
Db 213 SFLASFLIGMAVTTG 227

RESULT 31
CBIB_SALTY STANDARD: PRT: 319 AA.
ID CBIB_SALTY
AC Q05600;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CBIB PROTEIN.
GN CBIB.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID:602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN:LT2;
RX MEDLINE:93273696; PubMed:8501034;
RA Roth J.R., Lawrence J.G., Rubenfield M., Kieffer-Higgins S.,
RA Church G.M.;
RA "Characterization of the cobalamin (vitamin B12) biosynthetic genes
RT of Salmonella typhimurium."
RL J. Bacteriol. 175:3303-3316(1993).
CC -1- FUNCTION: INVOLVED IN THE CONVERSION OF COBAYIC ACID TO
CC COBANIMIDE. ADDITION OF AMINOPROPANOL ON THE F CARBOXYLIC
CC GROUP.
CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE COB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L12006; AAA27253.1; -.
DR StGene: SG10035; cbib.
KW Cobalamin biosynthesis; Transmembrane.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 56 76 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
SQ SEQUENCE 319 AA: 35370 MW: 540FDF93E0796214 CRC64:

Query Match 46.0% Score 40; DB 1; Length 319;
Best Local Similarity 53.8% Pred. No. 83;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 LVVFFLLMGVTWG 16
Db 59 VVVVGATVGVAMG 71

RESULT 32
YHHT_ECOLI STANDARD: PRT: 349 AA.
ID YHHT_ECOLI
AC P37622; P76700;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN YHHT.
GN YHHT OR B3474.
OS Escherichia coli.
OC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae:
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN-K12 / MG1655.
RX MEDLINE=94316500; PubMed=8041620;
RA Soita H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE UPF0118 (PERR) FAMILY. STRONG, TO
CC H. INFLUENZAE H10338.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U00039; CAB34657.1; ALT_INIT.
DR EMBL: AE000423; AAC76499.1; ALT_INIT.
DR EcoGene: EG12220; yhht.
DR InterPro: IPR002549; UPF0118.
DR Pfam: PF01594; UPF0118; 1.
DR KMW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 63 83 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 209 229 POTENTIAL.
FT TRANSMEM 235 255 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
SQ SEQUENCE 349 AA; 38522 MW; 275ADBADA6F58257 CRC64;

Query Match 46.0%; Score 40; DB 1; Length 349;
Best Local Similarity .64.3%; Pred. No. 89; 1; Indels 2; Gaps 1;
Matches 9; Conservative 2; Mismatches 1;

QY 1 MSMLVF--LLMG 12
| : | | | : | : | |
DB 290 MSTMVVPLSLIMG 303

RESULT 33
OPSD_CHELB STANDARD: PRT: 353 AA.
AC 09YGZ8:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RHODOPSIN.
GN RHO.
OS Cheilus labrosus (Thicklip grey mullet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Mugilomorpha; Mugilidae;
OC Cheilus.
OX NCBI_TaxID=48171;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina.
RA Archer S.N., Hirano J.;
RT "Comparative analysis of opsins in Mediterranean coastal fish.";

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Y18669; CAI77251.1; -;
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR InterPro: IPR001760; Opsin.
DR Pfam: PF00001; 7tm.1; 1.
DR PRINTS: PR00237; GPCR_RHODOPSN.
DR PRINTS: PR00238; OPSIN.
DR PRINTS: PR00579; RHODOPSIN.
DR PROSITE: PS00237; G-PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS50262; G-PROTEIN_RECEP_F1_2; 1.
DR KMW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
FT DOMAIN 1 36
FT TRANSMEM 37 61 1 (POTENTIAL).
FT DOMAIN 62 73 CYTOPLASMIC.
FT TRANSMEM 74 98 2 (POTENTIAL).
FT DOMAIN 99 113 EXTRACELLULAR.
FT TRANSMEM 114 133 3 (POTENTIAL).
FT DOMAIN 134 152 CYTOPLASMIC.
FT TRANSMEM 153 176 4 (POTENTIAL).
FT DOMAIN 177 202 5 (POTENTIAL).
FT TRANSMEM 203 230 6 (POTENTIAL).
FT DOMAIN 231 252 CYTOPLASMIC.
FT TRANSMEM 253 276 7 (POTENTIAL).
FT DOMAIN 277 284 EXTRACELLULAR.
FT TRANSMEM 285 309 7 (POTENTIAL).
FT DOMAIN 310 353 CYTOPLASMIC.
FT CARBOHYD 2
FT CARBOHYD 15 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT BINDING 296 296 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 322 322 RETINAL CHROMOPHORE (BY SIMILARITY).
FT DISULFID 110 187 PALMITATE (BY SIMILARITY).
FT SEQUENCE 353 AA; 39505 MW; 8C4B6F4847B3C632 CRC64;

Query Match 46.0%; Score 40; DB 1; Length 353;
Best Local Similarity 47.1%; Pred. No. 90;
Matches 8; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

QY 3 MLVFLLM----GVTV 15
| : | | | : | | | |
DB 257 MVVAFVLCWCPYGVAVM 273

RESULT 34
OPSD_LITMO STANDARD: PRT: 353 AA.
AC 09YH00:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RHODOPSIN.

GN RHO.
 OS Lithognathus mormyrus (Striped seabream).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Perciformes; Percoidae;
 OC Sparidae; Lithognathus.
 OX NCBI_TaxID:50591;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE:Retina;
 RA Archer S.N., Hirano J.;
 RT "Comparative analysis of opsins in Mediterranean coastal fish."
 RL Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 CC LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
 CC VISION IN DIM LIGHT.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 CC BE PHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: Y18667; CAA7249.1; -
 CC DR InterPro: IPR000276; GPCR_Rhodopsn.
 CC DR InterPro: IPR001760; Opsin.
 CC DR Pfam: PF00001; 7tm_1; 1.
 CC DR PRINTS: PR00237; GPCR_RHODOPSIN.
 CC DR PRINTS: PR00238; OPSIN.
 CC DR PRINTS: PS00579; RHODOPSIN.
 CC DR PROSITE: PS00237; G-PROTEIN_RECIP_F1_1; 1.
 CC DR PROSITE: PS00262; G-PROTEIN_RECIP_F1_2; 1.
 CC DR PROSITE: PS00238; OPSIN; 1.
 CC DR PROSITE: PS00238; OPSIN; 1.
 CC KM Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 CC Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
 CC -----
 CC FT DOMAIN 1 36 EXTRACELLULAR.
 CC FT TRANSMEM 37 61 1 (POTENTIAL).
 CC FT DOMAIN 62 73 CYTOPLASMIC.
 CC FT TRANSMEM 74 98 2 (POTENTIAL).
 CC FT DOMAIN 99 113 EXTRACELLULAR.
 CC FT TRANSMEM 114 133 3 (POTENTIAL).
 CC FT DOMAIN 134 152 CYTOPLASMIC.
 CC FT TRANSMEM 153 176 4 (POTENTIAL).
 CC FT DOMAIN 177 202 5 (POTENTIAL).
 CC FT TRANSMEM 203 230 6 (POTENTIAL).
 CC FT DOMAIN 231 252 CYTOPLASMIC.
 CC FT TRANSMEM 253 276 7 (POTENTIAL).
 CC FT DOMAIN 277 284 EXTRACELLULAR.
 CC FT TRANSMEM 285 309 7 (POTENTIAL).
 CC FT DOMAIN 310 353 CYTOPLASMIC.
 CC FT CARBOHYD 2 61 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC FT CARBOHYD 15 73 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC FT CARBOHYD 200 200 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC FT BINDING 296 296 RETINAL CHROMOPHORE (BY SIMILARITY).
 CC FT LIPID 322 322 PALMITATE (BY SIMILARITY).
 CC FT DISULFID 110 187 BY SIMILARITY.
 CC SQ SEQUENCE 353 AA; 39525 MW; 38FB29AAD00EEFF87 CRC64;

Query Match 46.0%; Score 40; DB 1; Length 353;
 Best Local Similarity 41.2%; Pred. No. 90;
 Matches 7; Conservative 3; Mismatches 3; Indels 4; Gaps 1;
 Oy 3 MLVFLLLM---GVTW 15

Db 257 WVIATFLICMCPYAGVAM 273
 ID OPD_LIZAU STANDARD: PRT: 353 AA.
 AC O9YGZ6;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE RHODOPSIN.
 GN RHO.
 OS Liza aurata (Golden grey mullet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Perciformes; Mugilomorphi; Mugilidae;
 OC Liza.
 OX NCBI_TaxID:48191;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE:Retina;
 RA Archer S.N., Hirano J.;
 RT "Comparative analysis of opsins in Mediterranean coastal fish."
 RL Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 CC LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
 CC VISION IN DIM LIGHT.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 CC BE PHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: Y18671; CAA7253.1; -
 CC DR InterPro: IPR000276; GPCR_Rhodopsn.
 CC DR InterPro: IPR001760; Opsin.
 CC DR Pfam: PF00001; 7tm_1; 1.
 CC DR PRINTS: PR00237; GPCR_RHODOPSIN.
 CC DR PRINTS: PR00238; OPSIN.
 CC DR PRINTS: PS00579; RHODOPSIN.
 CC DR PROSITE: PS00237; G-PROTEIN_RECIP_F1_1; 1.
 CC DR PROSITE: PS00262; G-PROTEIN_RECIP_F1_2; 1.
 CC DR PROSITE: PS00238; OPSIN; 1.
 CC KM Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 CC Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
 CC -----
 CC FT DOMAIN 1 36 EXTRACELLULAR.
 CC FT TRANSMEM 37 61 1 (POTENTIAL).
 CC FT DOMAIN 62 73 CYTOPLASMIC.
 CC FT TRANSMEM 74 98 2 (POTENTIAL).
 CC FT DOMAIN 99 113 EXTRACELLULAR.
 CC FT TRANSMEM 114 133 3 (POTENTIAL).
 CC FT DOMAIN 134 152 CYTOPLASMIC.
 CC FT TRANSMEM 153 176 4 (POTENTIAL).
 CC FT DOMAIN 177 202 5 (POTENTIAL).
 CC FT TRANSMEM 203 230 6 (POTENTIAL).
 CC FT DOMAIN 231 252 CYTOPLASMIC.
 CC FT TRANSMEM 253 276 7 (POTENTIAL).
 CC FT DOMAIN 277 284 EXTRACELLULAR.
 CC FT TRANSMEM 285 309 7 (POTENTIAL).
 CC FT DOMAIN 310 353 CYTOPLASMIC.
 CC FT CARBOHYD 2 61 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC FT CARBOHYD 15 73 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC FT CARBOHYD 200 200 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC FT BINDING 296 296 RETINAL CHROMOPHORE (BY SIMILARITY).
 CC FT LIPID 322 322 PALMITATE (BY SIMILARITY).
 CC FT DISULFID 110 187 BY SIMILARITY.
 CC SQ SEQUENCE 353 AA; 39525 MW; 38FB29AAD00EEFF87 CRC64;

FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT BINDING 296 296 RETINAL CHROMOPHORE (BY SIMILARITY).
FT LIPID 322 322 PALMITATE (BY SIMILARITY).
FT DISULFID 110 187 BY SIMILARITY.
SQ SEQUENCE 353 AA; 39530 MW; D07AC0DC0B55EF9F CRC64;

Query Match 46.0%; Score 40; DB 1; Length 353;
Best Local Similarity 47.18; Pred. No. 90;
Matches 8; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

OY 3 MLVFFLLM---GVTW 15
1:1 11:1 11:1
Db 257 MVAFILCMCPYAGVAM 273

RESULT 36
OPSD_SOLS0
ID OPSD_SOLS0 STANDARD; PRT; 353 AA.
AC 09YG25;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE RHODOPSIN.
GN RHO.
OS Solea solea (Common sole).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Soleoidei; Soleidae; Solea.
OX NCBI_Taxid=90069;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Archer S.N., Hirano J.;
RT *Comparative analysis of opsins in Mediterranean coastal fish.*;
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
CC VISION IN DIM LIGHT.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Y18672; CAA77254.1; -;
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR InterPro: IPR001760; Opsin.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCR_RHODOPSIN.
DR PRINTS: PR00238; OPSIN.
DR PRINTS: PR00579; RHODOPSIN.
DR PROSITE: PS00237; G-PROTEIN_RECEPTOR_FL_1; 1.
DR PROSITE: PS00262; G-PROTEIN_RECEPTOR_FL_2; 1.
KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.
FT DOMAIN 1 36 EXTRACELLULAR.
FT TRANSMEM 37 61 1 (POTENTIAL).
FT DOMAIN 62 73 CYTOPLASMIC.
FT TRANSMEM 74 98 2 (POTENTIAL).
FT DOMAIN 99 113 EXTRACELLULAR.

FT TRANSMEM 114 133 3 (POTENTIAL).
FT DOMAIN 134 152 CYTOPLASMIC.
FT TRANSMEM 153 176 4 (POTENTIAL).
FT DOMAIN 177 202 EXTRACELLULAR.
FT TRANSMEM 203 230 5 (POTENTIAL).
FT DOMAIN 231 252 CYTOPLASMIC.
FT TRANSMEM 253 276 6 (POTENTIAL).
FT DOMAIN 277 284 EXTRACELLULAR.
FT TRANSMEM 285 309 7 (POTENTIAL).
FT DOMAIN 310 353 CYTOPLASMIC.
FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT BINDING 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT BINDING 296 296 RETINAL CHROMOPHORE (BY SIMILARITY).
FT LIPID 322 322 PALMITATE (BY SIMILARITY).
FT DISULFID 110 187 BY SIMILARITY.
SQ SEQUENCE 353 AA; 39616 MW; B22570A75EB20FA CRC64;

Query Match 46.0%; Score 40; DB 1; Length 353;
Best Local Similarity 41.28; Pred. No. 90;
Matches 7; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

OY 3 MLVFFLLM---GVTW 15
1:1 11:1 11:1
Db 257 MVAFILCMCPYAGVAM 273

RESULT 37
OPSU_BRARE
ID OPSU_BRARE STANDARD; PRT; 354 AA.
ID P35359;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ULTRAVIOLET-SENSITIVE OPSIN (ULTRAVIOLET CONE PHOTORECEPTOR PIGMENT)
DE (ZFO02).
GN ZFO2.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA MEDLINE=93317613; PubMed=8327475;
RA Robinson J., Schmitt E.A., Harosi F.I., Reece R.J., Dowling J.E.;
RT "Zebrafish ultraviolet visual pigment: absorption spectrum, sequence,
RT and localization.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6009-6012(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96244282; PubMed=8924413;
RA Robinson J., Schmitt E.A., Dowling J.E.;
RT "Temporal and spatial patterns of opsin gene expression in zebrafish
RT (Danio rerio)." 12:895-906(1995).
RL Vis. Neurosci. 12:895-906(1995).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: THE COLOR PIGMENTS ARE FOUND IN THE CONE
CC PHOTORECEPTOR CELLS.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED (BY SIMILARITY).
CC -1- MISCELLANEOUS: THIS OPSIN HAS AN ABSORPTION MAXIMUM AT 360 NM.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its


```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M69160: AAA25868.1: -.
DR PDB: 1BOG: 18-MAY-99.
KW Lyase; 3D-structure.
SQ SEQUENCE 451 AA; 49572 MW; 04CDB563BCB1C6EA CRC64;

Query Match 46.0%; Score 40; DB 1; Length 451;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 11 MGVTTG 16
Db 338 MGLTNG 343

RESULT 40
ID GUDH_BACSU STANDARD; PRT: 455 AA.
AC PA2238:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE GLUCARATE DEHYDRATASE (EC 4.2.1.40) (GDH) (GLUCD).
GN GUDH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95219079; PubMed=7704254;
RA Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.;
RT "Determination of a 21548 bp nucleotide sequence around the 24
RT degrees region of the Bacillus subtilis chromosome.";
RL Microbiology 141:269-275(1995).
CC -1- FUNCTION: CATALYZES THE DEHYDRATION OF GLUCARATE TO 5-KETO-4-
CC DEOXY-D-GLUCARATE (5-KDGLUC) (BT SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-GLUCARATE -> 5-DEHYDRO-4-DEOXY-D-GLUCARATE +
CC H(2)O..
CC -1- PATHWAY: FIRST STEP IN GLUCARATE CATABOLISM.
CC -1- SIMILARITY: BELONGS TO THE MANDELATE RACEMASE / MUCONATE
CC LACTONIZING ENZYME FAMILY. GLUCD SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D30808: BAA06470.1: -.
DR EMBL: 299105: CAB12043.1: -.
DR Subtilist; Bg11161; guld.
KW Lyase; Complete proteome.
SQ SEQUENCE 455 AA; 50782 MW; 3238486007698C2A CRC64;

Query Match 46.0%; Score 40; DB 1; Length 455;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 11 MGVTTG 16
Db 342 MGLTNG 347
```

```
RESULT 41
ID FHUB_ECOLI STANDARD; PRT: 660 AA.
AC P06972: P77372;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FERRICHRONE TRANSPORT SYSTEM PERMEASE PROTEIN FHUB (FERRICHRONE UPTAKE
DE PROTEIN FHUB).
GN FHUB OR B0153.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=87014116; PubMed=3020380;
RA Koester W., Braun V.;
RT "Iron hydroxamate transport of Escherichia coli: nucleotide sequence
RT of the fhb gene and identification of the protein.";
RL Mol. Gen. Genet. 204:435-442(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S. Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federpiet N., Hyman R., Kalman S., Komp C., Kurd O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oetner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-23 FROM N.A.
RX MEDLINE=88038363; PubMed=2823072;
RA Burkhardt R., Braun V.;
RT "Nucleotide sequence of the fhuc and fhud genes involved in iron
RT (III) hydroxamate transport: domains in fhuc homologous to
RT ATP-binding proteins.";
RL Mol. Gen. Genet. 209:49-55(1987).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INVOLVED IN THE HIGH-
CC AFFINITY TRANSPORT OF IRON(III)-FERRICHRONE INTO THE E. COLI
CC CELL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE FECCD
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04319: CAA27852.1: -.
CC
```

DR EMBL: X04319; CAA27853.1; ALT_INIT.
DR EMBL: D26562; CAB20273.1; -
DR EMBL: AE000124; AAC73264.1; -
DR EMBL: U70214; AAB08583.1; -
DR EMBL: X05810; CAA29256.1; -
DR PIR: S07318; S07318.
DR PIR: S45222; S45222.
DR EcoGene: EG10303; fnuB.
DR InterPro: IPR000522; RecCD.
DR Pfam: PF01032; RecCD_family; 2.
DR ProDom: PD001557; RecCD; 2.
DR Transport; Iton Transport; Transmembrane; Inner membrane;
KW Complete Proteome.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 63 83 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 197 217 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 277 297 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
FT TRANSMEM 348 368 POTENTIAL.
FT TRANSMEM 391 411 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
FT TRANSMEM 447 467 POTENTIAL.
FT TRANSMEM 479 499 POTENTIAL.
FT TRANSMEM 528 548 POTENTIAL.
FT TRANSMEM 567 587 POTENTIAL.
FT TRANSMEM 607 627 POTENTIAL.
FT TRANSMEM 635 655 POTENTIAL.
FT TRANSMEM 11 11 L -> V (IN REF. 5).
FT CONFLICT 311 311 T -> I (IN REF. 1 AND 2).
FT CONFLICT 364 364 MISSING (IN REF. 1 AND 2).
SO SEQUENCE 660 AA; 70422 MM; 427313E7B421B0BC CRC64;

Query Match 46.0%; Score 40; DB 1; Length 660;
Best Local Similarity 58.3%; Pred. No. 1.5e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 VVFLKMGVTWG 16
DB 127 VVGLIVFGVANG 138

RESULT 42
ID Y281_BUCAI STANDARD: PRT: 307 AA.
AC P57368;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN BU281.
GN BU281.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
synthetic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID:118099;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-TOKYO 1998;
RX MEDLINE-20445173; PubMed-10993077;
RX Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE EAMA TRANSPORTER FAMILY. STRONG, TO
CC S. TYPHIMORIUM PACO.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP001118; BAB1291.1; -
DR InterPro: IPR000620; DUF6.
DR InterPro: IPR001230; Prenyltn.
DR Pfam: PF00892; DUF6; 2.
DR Hypothetical protein; Transport; Transmembrane; Complete proteome.
KW TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 41 61 POTENTIAL.
FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 95 115 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 146 166 POTENTIAL.
FT TRANSMEM 174 194 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 239 259 POTENTIAL.
FT TRANSMEM 267 287 POTENTIAL.
SO SEQUENCE 307 AA; 35006 MM; 5D1851AECF6A6CE CRC64;

Query Match 45.4%; Score 39.5; DB 1; Length 307;
Best Local Similarity 44.4%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 3; Indels 5; Gaps 1;

QY 3 MLVFL-----LMLCVTW 15
DB 4 ILVILFSLVSLTWCTW 21

RESULT 43
ID UHPT_CHLPN STANDARD: PRT: 455 AA.
AC Q927N9; Q9J019;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE HEXOSE PHOSPHATE TRANSPORT PROTEIN.
GN CPN0665 OR CP0082.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID:83558;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-CWL029;
RX MEDLINE-99206606; PubMed-10192388;
RX Kaiman S., Mitchell W., Marathe R., Lamme C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-AR39;
RX MEDLINE-20150255; PubMed-10684935;
RX Reed T.D., Brunham R.C., Shen C., Gali S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Ueberback T., Berry K., Baas S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Salzberg S.L.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-J138;
RX MEDLINE-20330349; PubMed-10871362;
RX Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA.";

RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- FUNCTION: TRANSPORT PROTEIN FOR SUGAR PHOSPHATE UPTAKE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001649; AAD18804.1; -.
CC EMBL: AE002171; AAF37968.1; -.
CC EMBL: AF002547; BAA98872.1; -.
CC TIGR: CP0082; -.
CC InterPro: IPR000849; GLPT_transprtr.
CC PROSITE: PS00942; GLPT_1.
CC -----
CC TRANSPORT: Sugar transport; Transmembrane; Complete proteome.
CC KM TRANSMEM 34 54 POTENTIAL.
CC FT TRANSMEM 70 90 POTENTIAL.
CC FT TRANSMEM 113 133 POTENTIAL.
CC FT TRANSMEM 161 181 POTENTIAL.
CC FT TRANSMEM 185 205 POTENTIAL.
CC FT TRANSMEM 257 277 POTENTIAL.
CC FT TRANSMEM 302 322 POTENTIAL.
CC FT TRANSMEM 331 351 POTENTIAL.
CC FT TRANSMEM 363 383 POTENTIAL.
CC FT TRANSMEM 394 414 POTENTIAL.
CC FT TRANSMEM 424 444 POTENTIAL.
CC SO SEQUENCE 455 AA; 51659 MW; F6404E2F1F4CA076 CRC64;

Query Match 45.4%; Score 39.5; DB 1; Length 455;
Best Local Similarity 40.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 5; Gaps 1;

Oy 2 SMLVFLMLMGVT-----WG 16
Db 119 SIVFLALMGLMGVFGWG 138

RESULT 44
UHPT_CHLTR STANDARD: PRT; 456 AA.
ID UHPT_CHLTR STANDARD: PRT; 456 AA.
AC 084548;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE HEXOSE PHOSPHATE TRANSPORT PROTEIN.
GN CT544.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusev R.L., Zhao O., Koonin E.V.,
RA Davis R.W.;
RT *Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.*;
RL Science 282:754-759(1998).
CC -1- FUNCTION: TRANSPORT PROTEIN FOR SUGAR PHOSPHATE UPTAKE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: AE001325; AAC68146.1; -.
DR InterPro: IPR000849; GLPT_transprtr.
DR InterPro: IPR003662; sub_transprtr.
DR Pfam: PF00083; Sugar_tr; 1.
CC PROSITE: PS00942; GLPT_1.
CC -----
CC TRANSPORT: Sugar transport; Transmembrane; Complete proteome.
CC KM TRANSMEM 34 54 POTENTIAL.
CC FT TRANSMEM 70 90 POTENTIAL.
CC FT TRANSMEM 113 133 POTENTIAL.
CC FT TRANSMEM 161 181 POTENTIAL.
CC FT TRANSMEM 185 205 POTENTIAL.
CC FT TRANSMEM 257 277 POTENTIAL.
CC FT TRANSMEM 302 322 POTENTIAL.
CC FT TRANSMEM 331 351 POTENTIAL.
CC FT TRANSMEM 363 383 POTENTIAL.
CC FT TRANSMEM 394 414 POTENTIAL.
CC FT TRANSMEM 421 441 POTENTIAL.
CC SO SEQUENCE 456 AA; 51729 MW; E443610D6B5B7E8F CRC64;

Query Match 45.4%; Score 39.5; DB 1; Length 456;
Best Local Similarity 35.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 5; Gaps 1;

Oy 2 SMLVFLMLMGVT-----WG 16
Db 119 STPLFLVFLMGVFGWG 138

RESULT 45
YXJN_BACSU STANDARD: PRT; 157 AA.
ID YXJN_BACSU STANDARD: PRT; 157 AA.
AC P55182;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 17.4 KDA PROTEIN IN GALE-PEPT INTERGENIC REGION.
GN YXJN.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97132636; PubMed=8978088;
RA Schroegel O., Krispin O., Allmannberger R.;
RT *Expression of a pept homologue from Bacillus subtilis.*;
RL FEMS Microbiol. Lett. 145:341-348(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT *Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the ltc and cel loci, and creation of a 177 kb contig
RT covering the gut-sacxy region.*;
RL Microbiology 142:3113-3123(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@lsb-slb.ch).

DR EMBL; X99339; CAA67715.1; -
DR EMBL; D83026; BAA11715.1; -
DR EMBL; Z99123; CAB15915.1; -
DR Subtilisin; B611897; YxjN.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 52 72 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
SQ SEQUENCE 157 AA; 17382 MW; DDDP8C768A6EC9F6 CRC64;

Query Match 44.8%; Score 39; DB 1; Length 157;
Best Local Similarity 50.0%; Pred. No. 65;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 MLVFLLMGVT 14
::: |||||:
Db 30 LIPLLLMGVS 41

Search completed: January 7, 2002, 16:50:36
Job time: 278 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2002, 16:51:26 ; Search time 31.95 Seconds
(without alignments)
73.251 Million cell updates/sec

Title: US-09-471-276-831_COPY_1_16
Perfect score: 87
Sequence: 1 MSMLVFLLMGVTWG 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 100 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	58.6	513	11	09EPH1 Oepl1 ratius norv
2	50	57.5	98	2	09RWN2 Oepl2 delinococcus
3	50	57.5	243	2	09WY46 Oepl46 thermotoga
4	50	57.5	251	2	09Z321 Oepl321 pseudomonas
5	50	57.5	443	2	09L051 Oepl051 streptomyces
6	48.5	55.7	228	2	09S96 Oepl96 staphylococ
7	48	55.2	471	2	P72122 P72122 pseudomonas
8	47	54.0	255	12	P87639 P87639 equine arte
9	47	54.0	255	12	P87640 P87640 equine arte
10	47	54.0	255	12	09YNU0 OeplNU0 equine arte
11	47	54.0	255	12	09YNT6 OeplNT6 equine arte
12	47	54.0	255	12	09OKG7 OeplKG7 equine arte
13	47	54.0	255	12	09OKG5 OeplKG5 equine arte
14	47	54.0	255	12	09OKG3 OeplKG3 equine arte
15	47	54.0	255	12	09OKG2 OeplKG2 equine arte
16	47	54.0	255	12	09PW24 OeplW24 equine arte
17	47	54.0	255	12	09PW4 OeplW4 equine arte
18	47	54.0	255	12	066483 Oepl6483 equine arte
19	47	54.0	255	12	0995W1 Oepl95W1 equine arte

20	47	54.0	255	12	0995W0 Oepl95W0 equine arte
21	47	54.0	255	12	0995V9 Oepl95V9 equine arte
22	47	54.0	255	12	098VL6 Oepl8VL6 equine arte
23	47	54.0	300	2	09HXB6 OeplHXB6 pseudomonas
24	47	54.0	389	2	09PGL7 OeplGL7 xylella fas
25	46	52.9	529	6	018736 Oepl8736 bos taurus
26	45.5	52.3	448	2	0914F8 Oepl14F8 pseudomonas
27	45	51.7	57	10	039398 Oepl39398 betula verr
28	45	51.7	177	2	P95489 P95489 pseudomonas
29	45	51.7	255	12	P87643 P87643 equine arte
30	45	51.7	255	12	066481 Oepl6481 equine arte
31	45	51.7	255	12	066495 Oepl6495 equine arte
32	45	51.7	255	12	066482 Oepl6482 equine arte
33	45	51.7	275	2	09RXV3 OeplRXV3 delinococcus
34	45	51.7	436	2	09X8H4 OeplX8H4 streptomyces
35	45	51.7	534	5	062194 Oepl2194 caenorhabdi
36	44.5	51.1	227	2	09PAE4 OeplPAE4 xylella fas
37	44	50.6	126	2	P73694 P73694 synechocyst
38	44	50.6	255	12	066484 Oepl6484 equine arte
39	44	50.6	255	12	0995W5 Oepl95W5 equine arte
40	44	50.6	280	4	09H3N1 OeplH3N1 homo sapien
41	44	50.6	287	4	09Y4T6 OeplY4T6 homo sapien
42	44	50.6	298	2	09PMCS OeplPMCS campylobact
43	44	50.6	379	10	049447 Oepl49447 arabidopsis
44	44	50.6	417	2	09FK7 OeplFK7 bacillus ha
45	44	50.6	471	2	09EVC6 OeplEVC6 actinobacil
46	44	50.6	486	10	09LXK0 OeplXK0 arabidopsis
47	44	50.6	494	5	023835 Oepl3835 drosophila
48	44	50.6	494	5	018345 Oepl8345 drosophila
49	44	50.6	507	2	099WU6 Oepl99WU6 staphylococ
50	44	50.6	688	10	09SHY2 OeplSHY2 arabidopsis
51	44	50.6	792	5	09W123 OeplW123 drosophila
52	44	50.6	1197	2	09DP9 OeplDP9 xylella fas
53	44	49.4	103	2	P73564 P73564 synechocyst
54	44	49.4	255	12	P87641 P87641 equine arte
55	44	49.4	255	12	P87645 P87645 equine arte
56	44	49.4	255	12	066490 Oepl6490 equine arte
57	44	49.4	255	12	066494 Oepl6494 equine arte
58	44	49.4	255	12	089854 Oepl89854 equine arte
59	44	49.4	255	12	088675 Oepl88675 equine arte
60	44	49.4	255	12	066480 Oepl66480 equine arte
61	44	49.4	255	12	066493 Oepl66493 equine arte
62	44	49.4	255	12	089626 Oepl89626 equine arte
63	44	49.4	255	12	0995W7 Oepl995W7 equine arte
64	44	49.4	255	12	0995W4 Oepl995W4 equine arte
65	44	49.4	305	2	045818 Oepl45818 chloroflexu
66	44	49.4	363	5	09VB97 OeplVB97 drosophila
67	44	49.4	402	5	024559 Oepl24559 drosophila
68	44	49.4	415	5	09VL50 OeplVL50 drosophila
69	44	49.4	428	5	09VLA9 OeplVLA9 drosophila
70	44	49.4	511	1	09H106 OeplH106 thermoplasma
71	44	49.4	621	10	092083 Oepl92083 arabidopsis
72	44	49.4	710	10	09SU01 Oepl9SU01 arabidopsis
73	44	49.4	3345	5	046074 Oepl46074 drosophila
74	44	49.4	3386	5	09NFD0 OeplNFD0 drosophila
75	44	49.4	3437	5	09W507 OeplW507 drosophila
76	42.5	48.9	218	2	09JUD4 OeplJUD4 neisseria m
77	42.5	48.9	218	2	09UR78 OeplUR78 neisseria m
78	42.5	48.9	307	2	09AGE0 OeplAGE0 yersinia ps
79	42	48.3	102	2	047295 Oepl47295 plasmid col
80	42	48.3	102	2	047298 Oepl47298 escherichia
81	42	48.3	103	2	09XDO5 OeplXDO5 desulfitoba
82	42	48.3	106	2	09Z4F6 OeplZ4F6 escherichia
83	42	48.3	121	5	09U1A3 OeplU1A3 leishmania
84	42	48.3	238	2	099V28 Oepl99V28 staphylococ
85	42	48.3	253	2	032400 Oepl32400 rhodospirillum
86	42	48.3	265	5	09U102 OeplU102 leishmania
87	42	48.3	338	1	09YB01 OeplYB01 aeropyrum p
88	42	48.3	406	2	09KDD0 OeplKDD0 bacillus ha
89	42	48.3	407	2	09RI96 OeplRI96 streptomyces
90	42	48.3	410	2	P72472 P72472 streptomyces
91	42	48.3	476	2	099R04 Oepl99R04 staphylococ
92	42	48.3	593	10	09FHT4 OeplFHT4 arabidopsis

93 42 48.3 615 5 Q9XTR0 Q9Xtk0 caenorhabdi
 94 41.5 47.7 113 2 Q9X9U7 Q9X9U7 streptomyce
 95 41.5 47.7 113 2 Q9X9S3 Q9X9S3 streptomyce
 96 41.5 47.7 793 2 Q9H2C9 Q9H2C9 pseudomonas
 97 41 47.1 255 12 P87644 P87644 equine arte
 98 41 47.1 255 12 P87646 P87646 equine arte
 99 41 47.1 255 12 Q9YV79 Q9YV79 equine arte
 100 41 47.1 255 12 Q9YV78 Q9YV78 equine arte

ALIGNMENTS

RESULT 1

Q9EPH1

ID Q9EPH1

AC Q9EPH1

DT 01-MAR-2001 (TRENBLREL. 16, Created)

DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)

DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)

DE PUTATIVE ALPHA 1B-GLYCOPROTEIN.

OS Rattus norvegicus (Rat)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

RN NCBI_Taxid:10116;

Query Match 58.6%; Score 51; DB 11; Length 513;
 Best Local Similarity 62.5%; Pred. No. 15;
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSMLVFLMLMGVTW 16
 11:1 11111 1
 DB 1 MSMLTVLLMLMGFTIG 16

RESULT 2

Q9RWN2

ID Q9RWN2

AC Q9RWN2

DT 01-MAY-2000 (TRENBLREL. 13, Created)

DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)

DT 01-MAR-2001 (TRENBLREL. 16, Last annotation update)

DE HYPOTHETICAL 10.7 KDA PROTEIN.

GN DR0634.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.

OX NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RA MEDLINE:20036896; PubMed:10567266;
 RA White O., Elisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.D., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Mofat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Yamathavan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.,
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans R1."
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001921; AAF10217.1; -.
 DR TIGR: DR0634; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 98 AA; 10696 MW; DA02F874C18BD6B5 CRC64;

Query Match 57.5%; Score 50; DB 2; Length 98;

Best Local Similarity 60.0%; Pred. No. 5;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSMLVFLMLMGVTW 15
 11:1 11111 1
 DB 1 MSMLVFLMLMGVTW 15

RESULT 3

Q9WY46

ID Q9WY46

AC Q9WY46

DT 01-NOV-1999 (TRENBLREL. 12, Created)

DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)

DT 01-MAY-2000 (TRENBLREL. 13, Last annotation update)

DE ABC TRANSPORTER, PERMEASE PROTEIN, CYSTW FAMILY.

GN TM0203.

OS Thermotoga maritima.

OC Bacteria; Thermotogales; Thermotoga.

OX NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

RN NCBI_Taxid:2336;

Query Match 57.5%; Score 50; DB 2; Length 243;
 Best Local Similarity 53.3%; Pred. No. 11;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSMLVFLMLMGVTW 15
 11:1 11111 1
 DB 97 VSMVLVFLMLMGVTW 111

RESULT 4

Q9Z3Z1

ID Q9Z3Z1

AC Q9Z3Z1

DT 01-MAY-1999 (TRENBLREL. 10, Created)

DT 01-MAY-1999 (TRENBLREL. 10, Last sequence update)

DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)

DE CMC PROTEIN (FRAGMENT).

GN CMC.

OS Pseudomonas putida.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OX NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

RN NCBI_Taxid:1299;

OC Pseudomonas.
 OX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=P8;
 RA Holtwick R., Keweloh H., Meinhardt F.;
 RT "cis/trans isomerase of unsaturated fatty acids of Pseudomonas putida
 P8: Evidence for a heme protein of the cytochrome c-type.";
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ131925; CAI0529.1; -;
 DR InterPro: IPR002541; CytoC.asn.
 DR InterPro: IPR003557; CytoC.biol_Ccmc.
 DR Pfam: PF01578; CytoC.asn.1.
 DR PRINTS: PR01386; CCMCHOGNSIS.
 FT INIT_MET 0
 FT NON_TER 251
 SQ SEQUENCE 251 AA; 28085 MW; 434636CCADBD34D CRC64;

Query Match 57.5%; Score 50; DB 2; Length 251;
 Best Local Similarity 56.2%; Pred. No. 11;
 Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSMLVFLLMGVTWG 16
 DB 27 LAALAVLLLVGVTWG 42

RESULT 5
 O91051 PRELIMINARY; PRT; 443 AA.
 AC O91051;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE PUTATIVE INTEGRAL MEMBRANE PROTEIN.
 GN SCES0.12.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.U., Harris D.;
 RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cereno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapalite D., Eichner A., Cullum J.,
 RA Kinshir H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL163672; CAB87334.1; -;
 SQ SEQUENCE 443 AA; 47464 MW; 1CEA47D3C453D4FC CRC64;

Query Match 57.5%; Score 50; DB 2; Length 443;
 Best Local Similarity 60.0%; Pred. No. 18;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 SMILVFLLMGVTWG 16
 DB 138 AFLVPEFLVLAFCVTWG 152

RESULT 6
 O99S96 PRELIMINARY; PRT; 228 AA.
 ID O99S96
 AC O99S96;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE SA1973 PROTEIN.
 GN SA1973.
 OS Staphylococcus aureus subsp. aureus N315.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=158879;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kuroda M., Ohba T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Iian J., Ito T., Kanamori M.,
 RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
 RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
 RA Hiraoka H., Kuwara S., Goto S., Yabuzaki J., Kanehisa M.,
 RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
 RA Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 RT aureus.";
 RL Lancet 357:1225-1240(2001).
 DR EMBL: AP003136; BAB43262.1; -;
 KW Complete proteome.
 SQ SEQUENCE 228 AA; 25807 MW; 03A01AE1402493A CRC64;

Query Match 55.7%; Score 48.5; DB 2; Length 228;
 Best Local Similarity 75.0%; Pred. No. 16;
 Matches 9; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 6 VELLMGVT-WG 16
 DB 127 VELLMGVTWG 138

RESULT 7
 P72122 PRELIMINARY; PRT; 471 AA.
 ID P72122
 AC P72122;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE OPRC OUTER MEMBRANE PROTEIN C.
 GN PA3789.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAO1;
 RX MEDLINE=96349120; PubMed=8760927;
 RA Yoneyama H., Nakae T.;
 RT "Protein C (OPRC) of the outer membrane of Pseudomonas aeruginosa is a
 RT copper-regulated channel protein.";
 RL Microbiology 142:2137-2144(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAO1;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wedman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).

DR EMBL: D28119; BAA05665.1; -
DR EMBL: AE004797; AAC07176.1; -
KW Complete proteome.
SO SEQUENCE 471 AA; 52232 MW; 021B71B12D86F9FC CRC64;

Query Match 55.2%; Score 48; DB 2; Length 471;
Best Local Similarity 57.1%; Pred. No. 35;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 2 MSMLVFFLLMGV 15
:|::|:|:|:|
Db 201 SLMLFMLSCMTW 214

RESULT 8
ID P87639 PRELIMINARY; PRT; 255 AA.
AC P87639;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE STRAIN 11958 ORF5.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID:11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-11958;
RA Archambault D., St-Laurent G., Lepage N.;
Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U46948; AAB50279.1; -
DR InterPro: IPR001332; Arteri_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; Arteri_glycop; 1.
DR ProDom: PD002371; Equine_virus_ORF5; 1.
SO SEQUENCE 255 AA; 28709 MW; 1312986A9A2793D5 CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFFLLMGV 13
:|::|:|:|:|
Db 2 LSMIVLFLMGV 14

RESULT 9
ID P87640 PRELIMINARY; PRT; 255 AA.
AC P87640;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE STRAIN 19933 ORF5.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID:11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-19933;
RA Archambault D., St-Laurent G., Lepage N.;
Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U46949; AAB50280.1; -
DR InterPro: IPR001332; Arteri_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; Arteri_glycop; 1.
DR ProDom: PD002371; Equine_virus_ORF5; 1.
SO SEQUENCE 255 AA; 28612 MW; 8DBA5667A1DA3C08 CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFFLLMGV 13
:|::|:|:|:|
Db 2 LSMIVLFLMGV 14

RESULT 10
ID O9YNT6 PRELIMINARY; PRT; 255 AA.
AC O9YNT6;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE STRAIN CA95G ORF5, ORF6 AND ORF57 GENES, COMPLETE CDS.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID:11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CA95G;
RA Balasuriya U.B.R., Hedges J.F., Patton J.F., Timoney P.J.,
McCollum W.H., Hurlinger P.J., MacLachlan N.J.;
"Sequence Analysis and Phylogenetic Characterization of Two Strains of
RT Equine Arteritis Virus Isolated from Persistently Infected Stallions
RT on the Same Ranch in California.";
RL J. Virol. 0:0-0(1996).
DR EMBL: U81023; AAD00488.1; -
DR InterPro: IPR001332; Arteri_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; Arteri_glycop; 1.
DR ProDom: PD002371; Equine_virus_ORF5; 1.
SO SEQUENCE 255 AA; 28695 MW; 099008C111552D05 CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFFLLMGV 13
:|::|:|:|:|
Db 2 LSMIVLFLMGV 14

RESULT 11
ID O9YNT6 PRELIMINARY; PRT; 255 AA.
AC O9YNT6;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CA95G ORF5, 6 AND 7 GENES, COMPLETE CDS.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID:11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN CA95GS;
RA Balasuriya U.B.R., Hedges J.F., Patton J.F., Timoney P.J.,
McCollum W.H., Hurlinger P.J., MacLachlan N.J.;
"Sequence Analysis and Phylogenetic Characterization of Two Strains of
RT Equine Arteritis Virus Isolated from Persistently Infected Stallions
RT on the Same Ranch in California.";
RL J. Virol. 0:0-0(1996).
DR EMBL: U81027; AAD00496.1; -
DR InterPro: IPR001332; Arteri_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; Arteri_glycop; 1.

DR Prodom: PD002371; Equine_virus_ORF5; 1.
SQ SEQUENCE 255 AA; 28661 MW; 0EB8D432115521F6 CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSMLVFLLMGV 13
DB 2 LSMIVLFLMGV 14

RESULT 12

090KG7 PRELIMINARY; PRT; 255 AA.
ID 090KG7
AC 090KG7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ORF5.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA;
RX MEDLINE=99394666; PubMed=10466790;
RA Balasuriya U.B.R., Hedges J.F., Nadler S.A., McCollum W.H.,
Timoney P.J., MacLachlan N.J.;
RT "Genetic Stability of Equine Arteritis Virus during Horizontal and
Vertical Transmission in an Outbreak of Equine Viral Arteritis.";
RT J. Gen. Virol. 80:1949-1958(1999).
DR EMBL: AF118772; AAD54698.1;
DR InterPro: IPR001332; Arteri_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; Arteri_glycop. 1.
DR Prodom: PD002371; Equine_virus_ORF5; 1.
SQ SEQUENCE 255 AA; 28918 MW; AB92FDBFD0D685B CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSMLVFLLMGV 13
DB 2 LSMIVLFLMGV 14

RESULT 13

090KG5 PRELIMINARY; PRT; 255 AA.
ID 090KG5
AC 090KG5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ORF5.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R2;
RX MEDLINE=99394666; PubMed=10466790;
RA Balasuriya U.B.R., Hedges J.F., Nadler S.A., McCollum W.H.,
Timoney P.J., MacLachlan N.J.;
RT "Genetic Stability of Equine Arteritis Virus during Horizontal and
Vertical Transmission in an Outbreak of Equine Viral Arteritis.";
RT J. Gen. Virol. 80:1949-1958(1999).
DR EMBL: AF118774; AAD54710.1;-

DR InterPro: IPR001332; Arteri_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; Arteri_glycop. 1.
DR Prodom: PD002371; Equine_virus_ORF5; 1.
SQ SEQUENCE 255 AA; 28730 MW; DAEBB71CD605C66B CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSMLVFLLMGV 13
DB 2 LSMIVLFLMGV 14

RESULT 14

090KG3 PRELIMINARY; PRT; 255 AA.
ID 090KG3
AC 090KG3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ORF5.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PI;
RX MEDLINE=99394666; PubMed=10466790;
RA Balasuriya U.B.R., Hedges J.F., Nadler S.A., McCollum W.H.,
Timoney P.J., MacLachlan N.J.;
RT "Genetic Stability of Equine Arteritis Virus during Horizontal and
Vertical Transmission in an Outbreak of Equine Viral Arteritis.";
RT J. Gen. Virol. 80:1949-1958(1999).
DR EMBL: AF118775; AAD54716.1;-
DR InterPro: IPR001332; Arteri_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; Arteri_glycop. 1.
DR Prodom: PD002371; Equine_virus_ORF5; 1.
SQ SEQUENCE 255 AA; 28767 MW; BB3C0F89525E37AB CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSMLVFLLMGV 13
DB 2 LSMIVLFLMGV 14

RESULT 15

090KG2 PRELIMINARY; PRT; 255 AA.
ID 090KG2
AC 090KG2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ORF5.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P2;
RX MEDLINE=99394666; PubMed=10466790;
RA Balasuriya U.B.R., Hedges J.F., Nadler S.A., McCollum W.H.,
Timoney P.J., MacLachlan N.J.;
RT "Genetic Stability of Equine Arteritis Virus during Horizontal and

```
RT Vertical Transmission in an Outbreak of Equine Viral Arteritis."
RL J. Gen. Virol. 80:1949-1958(1999).
DR EMBL: AF118776; AAD54722.1; -.
DR InterPro: IPR001332; Arter1_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; Arter1_glycop. 1.
DR ProDom: PD002371; Equine_virus_ORF5. 1.
SQ SEQUENCE 255 AA; 28798 MW; 479BC0A1E49116F3 CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSMLVFLLMGV 13
Db 2 LSMIVLFLLMGV 14

RESULT 16
Q9PWZ4 PRELIMINARY; PRT; 255 AA.
ID Q9PWZ4;
AC Q9PWZ4;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE ORF5.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID:11047;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN:43, A1 AND A2;
RX MEDLINE:99394666; PubMed:10466790;
RA Balasuriya U.B.R., Hedges J.F., Nadler S.A., McColium W.H.,
RA Timoney P.J., MacLachlan N.J.;
*Genetic Stability of Equine Arteritis Virus during Horizontal and
RT Vertical Transmission in an Outbreak of Equine Viral Arteritis.";
RL J. Gen. Virol. 80:1949-1958(1999).
DR EMBL: AF118771; AAD54692.1; -.
DR EMBL: AF118769; AAD54680.1; -.
DR EMBL: AF118770; AAD54686.1; -.
DR InterPro: IPR001332; Arter1_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; Arter1_glycop. 1.
DR ProDom: PD002371; Equine_virus_ORF5. 1.
SQ SEQUENCE 255 AA; 28739 MW; 9A76A527041E18CA CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSMLVFLLMGV 13
Db 2 LSMIVLFLLMGV 14

RESULT 17
Q9PWV4 PRELIMINARY; PRT; 255 AA.
ID Q9PWV4;
AC Q9PWV4;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE ORF5.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID:11047;
RN 11
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-RQ, R1, AND BT-PA96;
RX MEDLINE:99394666; PubMed:10466790;
RA Balasuriya U.B.R., Hedges J.F., Nadler S.A., McColium W.H.,
RA Timoney P.J., MacLachlan N.J.;
RT "Genetic Stability of Equine Arteritis Virus during Horizontal and
RT Vertical Transmission in an Outbreak of Equine Viral Arteritis.";
RL J. Gen. Virol. 80:1949-1958(1999).
DR EMBL: AF118782; AAD54758.1; -.
DR EMBL: AF118773; AAD54704.1; -.
DR EMBL: AF118781; AAD54752.1; -.
DR InterPro: IPR001332; Arter1_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; Arter1_glycop. 1.
DR ProDom: PD002371; Equine_virus_ORF5. 1.
SQ SEQUENCE 255 AA; 28725 MW; DB72F01773B4AEC CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSMLVFLLMGV 13
Db 2 LSMIVLFLLMGV 14

RESULT 18
Q66483 PRELIMINARY; PRT; 255 AA.
ID Q66483; Q9YNU3;
AC Q66483;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE EAV CAN86 ORF5 (FIELD STRAIN CAN86 ORF5 5, 6 AND 7 GENES, COMPLETE
DE CDS).
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID:11047;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-EAV CAN86 (FIELD ISOLATE);
RX MEDLINE:96130216; PubMed:8553578;
RA Balasuriya U.B., Timoney P.J., McColium W.H., MacLachlan N.J.;
RT "Phylogenetic analysis of open reading frame 5 of field isolates of
RT equine arteritis virus and identification of conserved and
RT nonconserved regions in the G1 envelope glycoprotein.";
RL Virology 214:690-697(1995).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-CAN86;
RA Balasuriya U.B.R., Hedges J.F., Patton J.F., Timoney P.J.,
RA McColium W.H., Hurlinger P.J., MacLachlan N.J.;
RT "Sequence Analysis and Phylogenetic Characterization of Two Strains of
RT Equine Arteritis Virus Isolated from Persistently Infected Stallions
RT on the Same Ranch in California.";
RL J. Virol. 0:0-0(1996).
DR EMBL: U38594; AAA92869.1; -.
DR EMBL: U81021; AAD00484.1; -.
DR InterPro: IPR001332; Arter1_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; Arter1_glycop. 1.
DR ProDom: PD002371; Equine_virus_ORF5. 1.
SQ SEQUENCE 255 AA; 28668 MW; 8D0665A012198E02 CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSMLVFLLMGV 13
Db 2 LSMIVLFLLMGV 14
```



```

RESULT 19
ID 0995W1 PRELIMINARY: PRT: 255 AA.
AC 0995W1;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GL PROTEIN.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_Taxid=11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9901008;
RA Larsen L.E., Storgaard T., Holm E.;
RT "Phylogenetic characterization of the GL protein of equine arteritis
RT virus isolated from semen of asymptotic stallions and fatal cases of
RT equine viral arteritis."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF247547; AAK37630.1; -
SQ SEQUENCE 255 AA; 28760 MW; 55DE905C49960EE4 CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFLLMGV 13
Db 2 LSMIVLFLMGV 14

RESULT 20
ID 0995W0 PRELIMINARY: PRT: 255 AA.
AC 0995W0;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GL PROTEIN.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_Taxid=11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9910583;
RA Larsen L.E., Storgaard T., Holm E.;
RT "Phylogenetic characterization of the GL protein of equine arteritis
RT virus isolated from semen of asymptotic stallions and fatal cases of
RT equine viral arteritis."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF247548; AAK37631.1; -
SQ SEQUENCE 255 AA; 28810 MW; 10D89E2056827CC2 CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFLLMGV 13
Db 2 LSMIVLFLMGV 14

RESULT 21
ID 0995V9 PRELIMINARY: PRT: 255 AA.
AC 0995V9;
DT 01-JUN-2001 (TReMBLrel. 17, Created)

```

```

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GL PROTEIN.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_Taxid=11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9910583;
RA Larsen L.E., Storgaard T., Holm E.;
RT "Phylogenetic characterization of the GL protein of equine arteritis
RT virus isolated from semen of asymptotic stallions and fatal cases of
RT equine viral arteritis."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF247549; AAK37632.1; -
SQ SEQUENCE 255 AA; 28825 MW; 51DCC0625A72EEB4 CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFLLMGV 13
Db 2 LSMIVLFLMGV 14

RESULT 22
ID 098VL6 PRELIMINARY: PRT: 255 AA.
AC 098VL6;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GL PROTEIN.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_Taxid=11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2344-9, AND 2344-8;
RA Larsen L.E., Storgaard T., Holm E.;
RT "Phylogenetic characterization of the GL protein of equine arteritis
RT virus isolated from semen of asymptotic stallions and fatal cases of
RT equine viral arteritis."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF247542; AAK37625.1; -
DR EMBL: AF247541; AAK37624.1; -
SQ SEQUENCE 255 AA; 28783 MW; 6B5DCCB795162EFF CRC64;

Query Match 54.0%; Score 47; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFLLMGV 13
Db 2 LSMIVLFLMGV 14

RESULT 23
ID 09HXB6 PRELIMINARY: PRT: 300 AA.
AC 09HXB6;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN PA3897.
DE PA3897.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

```

```
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Slover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Smith K.A., Spencer D.H., Hancock R.E.W., Lory S., Olson M.V.;
RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT *Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.*;
RL Nature 406:959-964(2000);
DR EMBL; AE004807; AAC07284.1; -.
DR InterPro; IPR000620; DUF6.
DR Pfam; PF00892; DUF6; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 300 AA; 32291 MW; 390033DEC9185B1C CRC64;

Query Match 54.0%; Score 47; DB 2; Length 300;
Best Local Similarity 40.0%; Pred. No. 33;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSMLVFFLLMGVTW 15
Db 3 LSVLITVLTLMGTTW 17

RESULT 24
O9PG17 PRELIMINARY; PRT: 389 AA.
AC O9PG17;
DT 01-OCT-2000 (TRENBLREL. 15, Created)
DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE TRANSPORT PROTEIN.
GN XP0281.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia J.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorfi H.,
RA Falcinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Freja J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitejima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.V., Madeira A.M.B.N., Madeira H.M.F., Marinho C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Montello-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nham A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Pelxoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Ouagigo R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
```

```
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000);
DR EMBL; AE003881; AAF83094.1; -.
DR InterPro; IPR002549; UPF0118.
DR Pfam; PF01594; UPF0118; 1.
KW Complete proteome.
SQ SEQUENCE 389 AA; 42813 MW; 7DA4518950158C7 CRC64;

Query Match 54.0%; Score 47; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 41;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 MLVFFLLMGVTWG 16
Db 344 MLVALLFLGSLWG 357

RESULT 25
O18736 PRELIMINARY; PRT: 529 AA.
AC O18736;
DT 01-JAN-1998 (TRENBLREL. 05, Created)
DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE UDP-GLUCURONOSYLTRANSFERASE.
OS Bos taurus (bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008677; BAA23359.1; -.
DR InterPro; IPR002213; UDPGT.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 529 AA; 60678 MW; 10EBCD920A415F76 CRC64;

Query Match 52.9%; Score 46; DB 6; Length 529;
Best Local Similarity 56.2%; Pred. No. 73;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 MSMLVFFLLMGVTWG 16
Db 8 VSVANFFLLMGFTLG 23

RESULT 26
O914F8 PRELIMINARY; PRT: 448 AA.
AC O914F8;
DT 01-MAR-2001 (TRENBLREL. 16, Created)
DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE TWO-COMPONENT SENSOR PHOQ.
GN PHOQ OR PA1180.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Slover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
```


DE EAV AUT68 ORF5.
 OS Equine arteritis virus (EAV).
 CC Viruses: ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Arteriviridae; Arterivirus.
 OX NCBI_TaxID=11047;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-EAV AUT68 (FIELD ISOLATE);
 RX MEDLINE:96130216; PubMed-853578;
 RA Balasuriya U.B., Timoney P.J., McColllum W.H., MacLachlan N.J.;
 RT *Phylogenetic analysis of open reading frame 5 of field isolates of
 RT equine arteritis virus and identification of conserved and
 RT nonconserved regions in the GL envelope glycoprotein.";
 RL Virology 214:690-697(1995).
 DR EMBL: U38592; AAA92897.1; -;
 DR InterPro: IPR001332; Arterl_glycop.
 DR InterPro: IPR003241; Equine_virus_ORF5.
 DR Pfam: PF00951; Arterl_glycop; 1.
 DR ProDom: PD002371; Equine_virus_ORF5; 1.
 SO SEQUENCE 255 AA; 28550 MW; 563D2501877035D5 CRC64;

Query Match 51.7%; Score 45; DB 12; Length 255;
 Best Local Similarity 61.5%; Pred. No. 54;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFFLLMGV 13
 Db 2 LSMIVLLSLMGV 14

RESULT 31
 O66495 PRELIMINARY: PRT: 255 AA.
 AC O66495.
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE EAV SW264 ORF5.
 OS Equine arteritis virus (EAV).
 CC Viruses: ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Arteriviridae; Arterivirus.
 OX NCBI_TaxID=11047;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-EAV SW264 (FIELD ISOLATES);
 RX MEDLINE:96130216; PubMed-853578;
 RA Balasuriya U.B., Timoney P.J., McColllum W.H., MacLachlan N.J.;
 RT *Phylogenetic analysis of open reading frame 5 of field isolates of
 RT equine arteritis virus and identification of conserved and
 RT nonconserved regions in the GL envelope glycoprotein.";
 RL Virology 214:690-697(1995).
 DR EMBL: U38609; AAA92914.1; -;
 DR InterPro: IPR001332; Arterl_glycop.
 DR InterPro: IPR003241; Equine_virus_ORF5.
 DR Pfam: PF00951; Arterl_glycop; 1.
 DR ProDom: PD002371; Equine_virus_ORF5; 1.
 SO SEQUENCE 255 AA; 28649 MW; DEC27A94838D8C10 CRC64;

Query Match 51.7%; Score 45; DB 12; Length 255;
 Best Local Similarity 61.5%; Pred. No. 54;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFFLLMGV 13
 Db 2 LSMIVLLSLMGV 14

RESULT 32
 O66482 PRELIMINARY: PRT: 255 AA.
 ID O66482
 AC O66482; G9YV80;

DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE EAV A287 ORF5.
 OS Equine arteritis virus (EAV).
 CC Viruses: ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Arteriviridae; Arterivirus.
 OX NCBI_TaxID=11047;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-EAV A287 (FIELD ISOLATE);
 RX MEDLINE:96130216; PubMed-853578;
 RA Balasuriya U.B., Timoney P.J., McColllum W.H., MacLachlan N.J.;
 RT *Phylogenetic analysis of open reading frame 5 of field isolates of
 RT equine arteritis virus and identification of conserved and
 RT nonconserved regions in the GL envelope glycoprotein.";
 RL Virology 214:690-697(1995).
 DR EMBL: U38593; AAA92898.1; -;
 DR InterPro: IPR001332; Arterl_glycop.
 DR InterPro: IPR003241; Equine_virus_ORF5.
 DR Pfam: PF00951; Arterl_glycop; 1.
 DR ProDom: PD002371; Equine_virus_ORF5; 1.
 KW Hypothetical protein.
 SO SEQUENCE 255 AA; 28697 MW; 290CFE42341FFD67 CRC64;

Query Match 51.7%; Score 45; DB 12; Length 255;
 Best Local Similarity 61.5%; Pred. No. 54;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFFLLMGV 13
 Db 2 LSMIVLLSLMGV 14

RESULT 33
 O9RXV3 PRELIMINARY: PRT: 275 AA.
 ID O9RXV3
 AC O9RXV3.
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN DR0203.
 OS Deinococcus radiodurans.
 CC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RL;
 RX MEDLINE:20036896; PubMed:10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT *Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans RL.";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001882; AAF09788.1; -.

DR TIGR: DR0203: -
 KW Complete proteome.
 SQ SEQUENCE 275 AA; 29820 MW; F17E96929ED9EAA1 CRC64;

Query Match
 Best Local Similarity 51.7%; Score 45; DB 2; Length 275;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFLMG 12
 : |||: |||
 DB 135 LPMVLVFLMG 146

RESULT 34
 O9XBH4 PRELIMINARY; PRT: 436 AA.

AC O9XBH4;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DE HYPOTHETICAL 47.7 KDA PROTEIN (FRAGMENT).
 GN SCE9.01.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kiser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinsahl H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL MOL. Microbiol. 21:77-96(1996).
 DR EMBL: AL049841; CAB42747.1; -
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 436
 SQ SEQUENCE 436 AA; 47711 MW; 09C531FCC16B6414 CRC64;

Query Match
 Best Local Similarity 51.7%; Score 45; DB 2; Length 436;
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 MLVFLMGVTWG 16
 : | | | | | | | | | |
 DB 179 VLAVSLLTGATWG 192

RESULT 35
 O62194 PRELIMINARY; PRT: 554 AA.
 AC O62194;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-MAY-2001 (TREMblrel. 16, Last annotation update)
 DE F31C3.6 PROTEIN.
 GN F31C3.6.

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cortage A.;
 RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9909613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL: Z92784; CAB07196.3; -
 SQ SEQUENCE 554 AA; 62566 MW; B3574BDDAF0050EC CRC64;

Query Match
 Best Local Similarity 51.7%; Score 45; DB 5; Length 554;
 Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 3 MLVFLMGVTWG 16
 : | : | : | : | |
 DB 270 VLIITLVNMG 283

RESULT 36
 O9PAB4 PRELIMINARY; PRT: 227 AA.

AC O9PAB4;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE HYPOTHETICAL PROTEIN XF2574.
 GN XF2574.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dotry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Klitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Martins E.A.B.N., Madeira A.M.F., Martino C.L.,
 RA Marques M.V., Martins E.A., Miyaki C.Y., Montelero-Vitolledo C.B.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Montelero-Vitolledo C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Savasak H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silva J.F., Silvestri M.L.Z., Siqueira M.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsunako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Melandri J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 Nature 406:151-159(2000).

DR EMBL: AE004064; AAF85371.1; -.
DR InterPro: IPR000326; PA_PTPase.
DR Pfam: PF01569; PAP2; 1.
DR SMART: SM00014; acidPPc; 1.
KW Hypothetical protein; Complete proteome.
SO SEQUENCE 227 AA; 25665 MW; 59C79A1CD9B4A43 CRC64;

Query Match 51.1%; Score 44.5; DB 2; Length 227;
Best Local Similarity 62.5%; Pred. No. 58;
Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

OY 1 MSMLVFL-LLMGVTV 15
DB 153 MAMTLVLLVLMVDV 168

RESULT 37
P73694 PRELIMINARY; PRT; 126 AA.

ID P73694
AC P73694;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYDROTHERMAL 14.2 KDA PROTEIN.
GN SL1698.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:97061201; PubMed:8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hosoya M., Sugita M., Sasamoto S., Kinura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-116(1996).
DR EMBL: D90908; BAA17741.1; -.
KW Hypothetical protein; Complete proteome.
SO SEQUENCE 126 AA; 14220 MW; 42CF97B378A9919 CRC64;

Query Match 50.6%; Score 44; DB 2; Length 126;
Best Local Similarity 57.1%; Pred. No. 41;
Matches 12; Conservative 1; Mismatches 2; Indels 6; Gaps 2;

OY 1 MSMLVFLLL- - - - - WGVTVG 16
DB 78 MSIL-PFLLLGALCMGVKMG 97

RESULT 38
O66484 PRELIMINARY; PRT; 255 AA.

ID O66484
AC O66484;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EAV ITA92 ORF5.
OS Equine arteritis virus (EAV).
OC Arteriviridae; Arterivirus.
NCBI_TaxID=11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EAV ITA92 (FIELD ISOLATE);
RX MEDLINE:96130216; PubMed:8553578;
RA Balasubra U.B., Timoney P.J., McCollum W.H., MacLachlan N.J.;
RT "Phylogenetic analysis of open reading frame 5 of field isolates of

RT equine arteritis virus and identification of conserved and
RT nonconserved regions in the GL envelope glycoprotein."
RT Virology 214:690-697(1995).
DR EMBL: U38598; AAA92903.1; -.
DR InterPro: IPR001332; ArterL_glycop.
DR InterPro: IPR003241; Equine_virus_ORF5.
DR Pfam: PF00951; ArterL_glycop; 1.
DR Prodom: PD002371; Equine_virus_ORF5; 1.
SO SEQUENCE 255 AA; 28731 MW; 215B7DC4FE7041E9 CRC64;

Query Match 50.6%; Score 44; DB 12; Length 255;
Best Local Similarity 61.5%; Pred. No. 74;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSMLVFLLLMGV 13
DB 2 LSMIVLLFVLMGV 14

RESULT 39
O995W5 PRELIMINARY; PRT; 255 AA.

ID O995W5
AC O995W5;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GL PROTEIN.
OS Equine arteritis virus (EAV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
NCBI_TaxID=11047;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HORNAR.
RA Larsen L.E., Storgaard T., Holm E.;
RT "Phylogenetic characterization of the GL protein of equine arteritis
RT virus isolated from semen of asymptomatic stallions and fatal cases of
RT equine viral arteritis."
RL Submitted (MAR-2000) to the EMBL/Genbank/DDJ databases.
DR EMBL: AF247539; AAK37622.1; -.
SO SEQUENCE 255 AA; 28756 MW; F1897CP4D3797ECB CRC64;

Query Match 50.6%; Score 44; DB 12; Length 255;
Best Local Similarity 53.8%; Pred. No. 74;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSMLVFLLLMGV 13
DB 2 LSMIVLLFVLMGV 14

RESULT 40
O9H3N1 PRELIMINARY; PRT; 280 AA.

ID O9H3N1
AC O9H3N1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE THIOREDOXIN-RELATED TRANSMEMBRANE PROTEIN.
GN TMX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eulheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsuo Y., Akiyama N., Nakamura H., Yodoi J., Noda M.,
RA Kizuka-Kondoh S.;
RT "Identification of a Novel Thioredoxin-related Transmembrane
RT Protein."
RL J. Biol. Chem. 0:0-0(2001).

DR EMBL: AB048246; BAB20629.1; -
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; thioRed.1
 DR PROSITE: PS00194; THIOREDOXIN.1.
 DR Redox-active center; Transmembrane.
 KW Redox-active center; Transmembrane.
 SQ SEQUENCE 280 AA; 31791 MW; A57E22481A97DE CRC64;

Query Match 50.6%; Score 44; DB 4; Length 280;
 Best Local Similarity 53.8%; Pred. No. 80;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 MLVFFLLMGVTW 15
 : | : | | | | :
 Db 11 LAVLVLLMGAPW 23

RESULT 41
 ID 09Y476 PRELIMINARY; PRT; 287 AA.
 AC 09Y476.
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHEICAL 32.5 KDA PROTEIN (FRAGMENT).
 GN DKF2P564E1962.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN.
 RA Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL080080; CAB45700.1; -
 DR HSSP: P07237; IMEX.
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; thioRed.1.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR PROSITE: PS00194; THIOREDOXIN.1.
 KW Hypothetical protein; Redox-active center.
 FT NON_TER
 RN [1]
 SO SEQUENCE 287 AA; 32535 MW; 733B66A812C6F7E5 CRC64;

Query Match 50.6%; Score 44; DB 4; Length 287;
 Best Local Similarity 53.8%; Pred. No. 82;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 MLVFFLLMGVTW 15
 : | : | | | | :
 Db 18 LAVLVLLMGAPW 30

RESULT 42
 ID 09PMC5 PRELIMINARY; PRT; 298 AA.
 AC 09PMC5.
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE INTEGRAL MEMBRANE PROTEIN.
 GN C31544C.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 CC Campylobacter.
 NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,

RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holroyd S.,
 RA Jags K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajadream M.A., Rutherford K.M., Van Vliet A.H.M.,
 RA Whitehead S., Barrall B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences."
 RL Nature 403:665-668(2000).
 DR EMBL: AL139078; CAB73960.1; -
 DR InterPro: IPR000620; DUF6.
 DR Pfam: PF00892; DUF6; 2.
 KW Complete proteome.
 SQ SEQUENCE 298 AA; 33709 MW; B656B2AEB5E101FF CRC64;

Query Match 50.6%; Score 44; DB 2; Length 298;
 Best Local Similarity 46.2%; Pred. No. 85;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 3 MLVFFLLMGVTW 15
 : | : | | | : |
 Db 10 LLIAMFLMGSSW 22

RESULT 43
 ID 049447 PRELIMINARY; PRT; 379 AA.
 AC 049447.
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ADP, ATP CARRIER-LIKE PROTEIN.
 GN P2109.60 OR A14G28390.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eudicots II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Koeltter P., Hempel S., Entlan K.-D., Hohelsel J.,
 RA Mewes H.W., Mayer K., Schueller C.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Rose M., Hempel S., Entlan K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC -I- INNER MEMBRANE (BY SIMILARITY).
 CC -I- SIMILARITY: TO THE MITOCHONDRIAL CARRIER FAMILY.
 DR EMBL: AL021749; CAA16877.1; -
 DR EMBL: AL161572; CAB79641.1; -
 DR Mendel: 28166; Atrah.1415;28166.
 DR InterPro: IPR001993; Mitoch_carrler.
 DR InterPro: IPR002067; Mit_carrler.
 DR Pfam: PF00153; mito_carr; 3.
 DR PRINTS: PR00926; MITOCARRIER.
 DR PROSITE: PS00215; MITOCH_CARRIER; 2.
 KW Inner membrane; Mitochondrion; Repeat; Transmembrane; Transport.
 SO SEQUENCE 379 AA; 40718 MW; C865FAD6968A2A9 CRC64;

Query Match 50.6%; Score 44; DB 10; Length 379;
 Best Local Similarity 60.0%; Pred. No. 1e+02;
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 SMLVFFLLMGVTWG 16
 : | : | | | | : |
 Db 283 SFLASFLMGVITIG 297

RESULT 44

09KFK7 ID 09KFK7 PRELIMINARY; PRT: 417 AA.
 AC 09KFK7; 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE BH0472 PROTEIN.
 CN BH0472.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID:86665;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE:20512582; PubMed :1058132.
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR InterPro: IPR001958; TCR_Tcra.
 DR PRINTS: PRO1035; TCRTETA.
 KM Complete proteome.
 SO SEQUENCE 417 AA; 45439 MW; 17C32402CB14EA9 CRC64;

Query Match

Best Local Similarity 50.6%; Score 44; DB 2; Length 417;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 VVFLLMGVTW 15
 I: ||||: |

Db 104 VLRLMGIAW 114

RESULT 45

09EYCG ID 09EYCG PRELIMINARY; PRT: 471 AA.
 AC 09EYCG; 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE PUTATIVE UNDECAPRENTL-PHOSPHATE GALACTOSEPHOSPHOTRANSFERASE.
 OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Actinobacillus.
 OX NCBI_TaxID:715;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=4074;
 RA Labrie J., Rioux S., Wade M.M., Champin F.R., Holman S.C.,
 RA Wilson W.W., Savoye C., Kobisch M., Galarneau C., Jacques M.;
 RT "Identification of genes involved in biosynthesis of Actinobacillus
 pleuropneumoniae serotype 1 O-antigen and biological properties of
 rough mutants.";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF329452; AAG45943.1; -.
 DR InterPro: IPR003362; Bact_transf.
 DR InterPro: IPR001993; Mitoch_carrier.
 DR Pfam: PF02397; Bact_transf.1
 DR PROSITE: PS00215; MITOCH_CARRIER; UNKNOWN_1.
 KW Transferase.
 SO SEQUENCE 471 AA; 55192 MW; C5B445ABD7B3513 CRC64;

OY 10 LMGVTWG 16
 |||||
 Db 110 LMGVTWG 116

Search completed: January 7, 2002, 16:51:30
 Job time: 277 sec

Query Match 50.6%; Score 44; DB 2; Length 471;
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

[illegible]

Chordata: Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (sites)
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsunura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A.,
Fujikawa,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hiro,M.,
Omori,Y., Oka,T., Suzuki,Y., Odayashi,M., Nishi,T., Shibahara,T.,
Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2175)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction: 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source 1..2175
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/clone:"LNC11861"
/clone_1lb:"LNC"
/tissue_type:"human lung"
/note:"Cloning vector pME18SFL3"
BASE COUNT 520 a 598 c 626 g 431 t
ORIGIN
alignment_scores:
Quality: 68.50 Length: 17
Ratio: 4.281 Gaps: 1
Percent Similarity: 94.118 Percent Identity: 82.353
alignment_block:
US-09-471-276-831_COPY_1_16 x AK027222/rev ..
Align seg 1/1 to reverse of: AK027222 from: 1 to: 2175
1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGly...ValThrTrpG1 16
|||||
1134 ATGTCCATGCTCGTCTTCTTCTGCTGCGTGAGCGCTCGGCG 1085
16 y 16
1084 T 1084
seq_name: gb_hcrg:AC012313
seq_documentation_block:
LOCUS AC012313 186451 bp DNA HTG 09-AUG-2001
DEFINITION Homo sapiens chromosome 19 clone CTD-2619J13, *** SEQUENCING IN
PROGRESS ***, 2 ordered pieces.
ACCESSION AC012313
VERSION AC012313.6 GI:15145671
KEYWORDS HTG: HTGS_PHA5E2; HTGS_ACTIVIERIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 186451)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 186451)

AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 9, 2001 this sequence version replaced g1:7711546.
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 181942: contig of 181942 bp in length
* 181943 182042: gap of unknown length
* 182043 186451: contig of 4409 bp in length.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 835906
Center Clone name: CTD-2619J13

Summary Statistics
Consensus quality: 184784 bases at least Q40
Consensus quality: 185394 bases at least Q30
Consensus quality: 185487 bases at least Q20
Estimated insert size: 194270; agarose-fp estimation
Estimated insert size: 185541; sum-of-contigs
estimation
Quality coverage: 0.95 in Q20 bases; agarose-fp
estimation
Quality coverage: 1 in Q20 bases; sum-of-contigs
estimation.
FEATURES
source 1..186451
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/chromosome:"19"
/clone:"CTD-2619J13"
BASE COUNT 44285 a 49190 c 49575 g 43301 t 100 others
ORIGIN
alignment_scores:
Quality: 68.50 Length: 17
Ratio: 4.281 Gaps: 1
Percent Similarity: 94.118 Percent Identity: 82.353
alignment_block:
US-09-471-276-831_COPY_1_16 x AC012313/rev ..
Align seg 1/1 to reverse of: AC012313 from: 1 to: 186451
1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGly...ValThrTrpG1 16
|||||
11643 ATGTCCATGCTCGTCTTCTTCTGCTGCGTGAGCGCTCGGCG 11594
16 y 16
11593 T 11593
seq_name: gb_hcrg:AC022955
seq_documentation_block:
LOCUS AC022955 159560 bp DNA HTG 16-MAR-2000
DEFINITION Homo sapiens clone RP11-153B10, WORKING DRAFT SEQUENCE, 44
unordered pieces.
ACCESSION AC022955
VERSION AC022955.3 GI:7249299

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 159560)
TITLE Btiren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens, clone RP11-153B10
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 159560)
Btiren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Bouslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Chepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fensator,J.,
Ferrela,P., Fitzhugh,N., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lander,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meunier,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisanu,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:7108171.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: 153_B-10
Center clone name: 153_B-10
Sequencing vector: M13; M77815: 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 132746 bases at least Q40
Consensus quality: 145820 bases at least Q30
Consensus quality: 151525 bases at least Q20
Insert size: 153000; agarose-ff
Insert size: 155260; sum-of-contigs
Quality coverage: 3.2 in Q20 bases; agarose-ff
Quality coverage: 3.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1065: contig of 1065 bp in length
* 1066 1165: gap of 100 bp
* 1166 2405: contig of 1240 bp in length
* 2406 2505: gap of 100 bp
* 2506 3712: contig of 1207 bp in length
* 3713 3812: gap of 100 bp
* 3813 5667: contig of 1855 bp in length
* 5668 5767: gap of 100 bp
* 5768 7537: contig of 1770 bp in length
* 7538 7637: gap of 100 bp
* 7638 9729: contig of 2092 bp in length
* 9730 9829: gap of 100 bp
* 9830 10997: contig of 1168 bp in length
* 10998 11097: gap of 100 bp
* 11098 12601: contig of 1504 bp in length
* 12602 12701: gap of 100 bp
* 12702 14331: contig of 1630 bp in length
* 14332 14431: gap of 100 bp
* 14432 15960: contig of 1529 bp in length
* 15961 16060: gap of 100 bp
* 16061 17698: contig of 1638 bp in length
* 17699 17798: gap of 100 bp
* 17799 20095: contig of 2297 bp in length
* 20096 20195: gap of 100 bp
* 20196 22180: contig of 1985 bp in length
* 22181 22280: gap of 100 bp
* 22281 24120: contig of 1840 bp in length
* 24121 24220: gap of 100 bp
* 24221 26233: contig of 2013 bp in length
* 26234 26333: gap of 100 bp
* 26334 28235: contig of 1902 bp in length
* 28236 28335: gap of 100 bp
* 28336 30862: contig of 2527 bp in length
* 30863 30962: gap of 100 bp
* 30963 34201: contig of 3239 bp in length
* 34202 34301: gap of 100 bp
* 34302 37474: contig of 3173 bp in length
* 37475 37574: gap of 100 bp
* 37575 40724: contig of 3150 bp in length
* 40725 40824: gap of 100 bp
* 40825 43315: contig of 2431 bp in length
* 43316 43415: gap of 100 bp
* 43416 46746: contig of 3331 bp in length
* 46747 46846: gap of 100 bp
* 46847 49873: contig of 3027 bp in length
* 49874 49973: gap of 100 bp
* 49974 53197: contig of 3224 bp in length
* 53198 53297: gap of 100 bp
* 53298 56360: contig of 3063 bp in length
* 56361 56460: gap of 100 bp
* 56461 60751: contig of 4291 bp in length
* 60752 60851: gap of 100 bp
* 60852 65004: contig of 4153 bp in length
* 65005 65104: gap of 100 bp
* 65105 69522: contig of 4418 bp in length
* 69523 69622: gap of 100 bp
* 69623 73289: contig of 3667 bp in length
* 73290 73389: gap of 100 bp
* 73390 76613: contig of 3224 bp in length
* 76614 76713: gap of 100 bp
* 76714 81153: contig of 4440 bp in length
* 81154 81253: gap of 100 bp
* 81254 85920: contig of 4667 bp in length
* 85921 86020: gap of 100 bp
* 86021 91545: contig of 5525 bp in length
* 91546 91645: gap of 100 bp
* 91646 96877: contig of 5232 bp in length
* 96878 96977: gap of 100 bp
* 96978 102129: contig of 5152 bp in length
* 102130 102229: gap of 100 bp
* 102230 107561: contig of 5332 bp in length
* 107562 107661: gap of 100 bp
* 107662 112452: contig of 4791 bp in length
* 112453 112552: gap of 100 bp
* 112553 117674: contig of 5122 bp in length
* 117675 117774: gap of 100 bp
* 117775 125917: contig of 8143 bp in length
* 125918 126017: gap of 100 bp
* 126018 132284: contig of 6267 bp in length
* 132285 132384: gap of 100 bp
* 132385 137619: contig of 5235 bp in length
* 137620 137719: gap of 100 bp
* 137720 144498: contig of 6779 bp in length
* 144499 144598: gap of 100 bp

* 144599 150489: contig of 5891 bp in length
* 150490 150589: gap of 100 bp
* 150590 159560: contig of 8971 bp in length.
Location/Qualifiers
1. 159560

FEATURES
SOURCE
1. 159560
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/clone:"RP11-153810"
/clone_1bp:"RP11-11 Human Male BAC"
1. 1065
misc_feature
/note:"assembly-fragment"
1166. 2405
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
2506. 3712
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
3813. 5667
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
5768. 7537
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
7638. 9729
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
9830. 10997
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
11098. 12601
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
12702. 14331
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
14432. 15960
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
16061. 17698
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
17799. 20095
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
20196. 22180
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
22281. 24120
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
24221. 26233
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
26334. 28235
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
28336. 30862
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
30963. 34201
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
34302. 37474
/note:"assembly-fragment"
misc_feature
/note:"assembly-fragment"
37575. 40724
/note:"assembly-fragment"

alignment_scores:
Quality: 55.00 Length: 16
Ratio: 4.231 Gaps: 0
Percent Similarity: 81.250 Percent Identity: 50.000

alignment_block:

US-09-471-276-831_COPY_1_16 x AC022955/rev ..

Align seg 1/1 to reverse of: AC022955 from: 1 to: 159560

1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrpGly 16

74421 TTGCGATGATTTCCCTTCTCTCTGTGTGGGCGCTCTCGGCGC 74374

seq_name: gb_hhg:AC022937

seq_documentation_block:

LOCUS AC022937 189679 bp DNA

DEFINITION Homo sapiens clone RP11-577C21, WORKING DRAFT SEQUENCE, 23
unordered pieces.

ACCESSION AC022937

VERSION AC022937.3 GI:7249297

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
1 (bases 1 to 189679)
Birren, B., Linton, L., Nussbaum, C. and Lander, E.
Unpublished
2 (bases 1 to 189679)
Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, K., Bede, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearrell, K., Dewar, K., Domino, M., Doyle, M., Fensholt, J.,
Ferrelle, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howard, J. C., Johnson, R., Jones, C., Kann, L., Karakas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McDurk, A., McKernan, K.,
McPheeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rotman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced g1.6997314.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 577C_21

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 174164 bases at least Q40

Consensus quality: 181501 bases at least Q30

Consensus quality: 184805 bases at least Q20

Insert size: 194000; agarose-fp

Insert size: 187479; sum-of-contrigs

Quality coverage: 3.8 in Q20 bases; agarose-fp

Quality coverage: 3.9 in Q20 bases; sum-of-contrigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1251: contig of 1251 bp in length
* 1252 1351: gap of 100 bp
* 1352 3428: contig of 2077 bp in length
* 3429 3528: gap of 100 bp
* 3529 5941: contig of 2413 bp in length
* 5942 6041: gap of 100 bp
* 6042 8291: contig of 2250 bp in length
* 8292 8391: gap of 100 bp
* 8392 11392: contig of 3001 bp in length
* 11393 11492: gap of 100 bp
* 11493 13568: contig of 2076 bp in length
* 13569 13668: gap of 100 bp

*	13669	16289:	contlg of	2591 bp	in	length
*	16660	16399:	gap of	100 bp		
*	16660	19204:	contlg of	2845 bp	in	length
*	19205	19304:	gap of	100 bp		
*	19305	23036:	contlg of	3732 bp	in	length
*	23037	23136:	gap of	100 bp		
*	23137	27387:	contlg of	4251 bp	in	length
*	27388	27487:	gap of	100 bp		
*	27488	32384:	contlg of	4897 bp	in	length
*	32385	32484:	gap of	100 bp		
*	32485	35750:	contlg of	5086 bp	in	length
*	35751	37670:	gap of	100 bp		
*	37671	44401:	contlg of	6731 bp	in	length
*	44402	44501:	gap of	100 bp		
*	44502	52775:	contlg of	8274 bp	in	length
*	52776	52875:	gap of	100 bp		
*	52876	60711:	contlg of	7836 bp	in	length
*	60712	60811:	gap of	100 bp		
*	60812	71744:	contlg of	10933 bp	in	length
*	71745	71844:	gap of	100 bp		
*	71845	84379:	contlg of	12535 bp	in	length
*	84380	84479:	gap of	100 bp		
*	84480	98175:	contlg of	13696 bp	in	length
*	98176	98275:	gap of	100 bp		
*	98276	112174:	contlg of	13899 bp	in	length
*	112175	112274:	gap of	100 bp		
*	112275	126338:	contlg of	14064 bp	in	length
*	126339	126438:	gap of	100 bp		
*	126439	135337:	contlg of	17099 bp	in	length
*	143638	143637:	gap of	100 bp		
*	143638	162270:	contlg of	18633 bp	in	length
*	162271	189679:	gap of	100 bp		
*	162371	189679:	contlg of	27309 bp	in	length

UNRES	Location/Qualifiers
source	1. 189679
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="RP11-577C21"
	1. clone11b=RPc1-11 Human Male BAC
	1. 1251
misc_feature	/note="assembly-fragment"
misc_feature	1352. .3428
misc_feature	/note="assembly-fragment"
misc_feature	3529. .5941
misc_feature	/note="assembly-fragment"
misc_feature	6042. .8291
misc_feature	/note="assembly-fragment"
misc_feature	8392. .11392
misc_feature	/note="assembly-fragment"
misc_feature	11493. .13568
misc_feature	/note="assembly-fragment"
misc_feature	13669. .16259
misc_feature	/note="assembly-fragment"
misc_feature	16360. .19204
misc_feature	/note="assembly-fragment"
misc_feature	19305. .23036
misc_feature	/note="assembly-fragment"
misc_feature	23137. .27387
misc_feature	/note="assembly-fragment"
misc_feature	27488. .32384
misc_feature	/note="assembly-fragment"
misc_feature	32485. .37570
misc_feature	/note="assembly-fragment"
misc_feature	37671. .44401
misc_feature	/note="assembly-fragment"
misc_feature	44502. .52775
misc_feature	/note="assembly-fragment"
misc_feature	52876. .60711
misc_feature	/note="assembly-fragment"
misc_feature	60812. .71744
misc_feature	/note="assembly-fragment"
misc_feature	71845. .84379
misc_feature	/note="assembly-fragment"

[illegible]

```

/product""ExpM protein"
/protein_id."CA11941.1"
/db_xref."GI:3319927"
/db_xref."SPTREMBL:O86196"
/translatiion.""MORRKTGIALGSCAAGMAHIGVFNALAEIGIDVDVAGCSIGA
LVGAAYATNNI,SMSEOWRGFGYWDVYLRMDLGMORCSLLRGDRVFNVAHLHTTQDI
EDCKIGKYVTTNLTGRELMEITGEGDLHQAMRASCMSQGLSPVRFNDYVLYGCAAVN
PVAPRLARAGADIVIVADLOHDASLKHODULSTKPTASAVDIDMEHVSDMSRIRE
RLRGRRSQAESSPTAMEIMSTSIQLLENLKMTRMAGDPPDVLYQPCQIATLDPH
RAQDAITAGYKAKAKMRDELPLATEL"
1370..1373
/genes""expM"
1370..2397
/genes""expM"
1381..2397
/genes""expM"
/function""regulation of virulence factors"
/codon_start..1
/transl_table..1
/evidence""experimental
/product""ExpM protein"
/protein_id."CA11942.1"
/db_xref."GI:3319928"
/db_xref."SPTREMBL:O86197"
/translatiion.""MEQPLAGKHLVLIDDAVFRSVLAGYLTISGASVREALNGDAL
SLEHHPDMLTCDLKMPPMGGEIEFLCLRLKSDPTIVISATSQMADIAKYLRGCV
QDYLKPIRDYTRLRDAVMSCLYDMFTSQLNELDQLOQMSLNSGPEAITKLADL
OPVPOQTILARCVRNYROLTFAEQGLVLDIAALSETELAFYCLDVTGCVNNGNLIAL
LRLRTFNGLEHLEHVDQDHRPLYPTLLKQVNNOLLRSALSDGREPLLVGYYHPIKOL
LLISAGLNLATLNVNEQGVYLVNSGVPLDTLEGAYLYNLQNLNYOCETIMQOIKNGGGRRLM
LTPF"
2593..2597
/genes""gailu"
2593..2797
/genes""gailu"
2604..2797
/genes""gailu"
/codon_start..1
/transl_table..1
/product""Gailu protein"
/protein_id."CA11943.1"
/db_xref."GI:3319929"
/db_xref.""SPTREMBL:O86198"
/translatiion.""MSTVNNKVKYKKAIVPACLGTRMLPATKATPKENLPLVDKPLIY
VNNCLAGINEILVTHSS"
VNNCLAGINEILVTHSS"

BASE COUNT      703 a      562 c      756 g      776 t

ORIGIN

alignment_scores:
    Quality:      54.00      length:      15
    Ratio:        4.154      Gaps:      0
    Percent Similarity:      86.667      Percent Identity:      60.000

alignment_block:
US-09-471-276-831_COPY_1_16 x ECJ224437 ..

Align seg 1/1 to: ECJ224437 from: 1 to: 2797

      2 SerMctLeuValValPheLeuLeuLeuTrpGlyValLThrTrpGly 16
      ::::::::::::::::::::::::::::
2150 GCCTTGATGTCGTCCTTCTTCCTTCGTCGCGGATATACATCGCG 2194

seq_name: gb_htg:AL136379

seq_documentation_block:
LOCUS      AL136379      142558 bp      DNA      HTG      18-JUL-2001
DEFINITION      Homo sapiens chromosome 1 clone RP5-881P19 map 942.11-42.3, ***
SEQUENCING IN PROGRESS ***, 6 unordered pieces.
ACCESSION      AL136379
VERSION      AL136379.18      GI:14970781
KEYWORDS      HTG; HTGS_PHASE1: HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_PULLTOP.
SOURCE      human.

```

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
<p> ORGANISM Homo sapiens Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo. 1 (bases 1 to 142558) Wray, P. Direct Submission Submitted (17-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk On Jul 19, 2001 this sequence version replaced gi:14529732. ----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquerry@sanger.ac.uk ----- Project Information Center project name: dj881p19 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 99% of reads Chemistry: Dye-primers: Big Dye; 0% of reads Consensus quality: 135267 bases at least Q40 Consensus quality: 140700 bases at least Q30 Consensus quality: 141500 bases at least Q20 Insert size: 142058; sum-of-contigs Insert size: 137842; 3.8% error; agarose-fp Quality coverage: 7.73x in Q20 bases; sum-of-contigs quality coverage: 9.52x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 6 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 2667: contig of 2667 bp in length * 2668 2767: gap of 100 bp * 2768 5694: contig of 2927 bp in length * 5695 5794: gap of 100 bp * 5795 7987: contig of 2193 bp in length * 7988 8087: gap of 100 bp * 8088 135903: contig of 127816 bp in length * 135904 136003: gap of 100 bp * 136004 139240: contig of 3237 bp in length * 139241 139341: gap of 100 bp * 139341 142558: contig of 3216 bp in length. Location/Qualifiers 1..142558 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /map="q42.11-q42.3" /clone="RP5-881p19" /clone_id="RPC1-5" 1..2667 /note="assembly_fragment:02282" 2768..5694 /note="assembly_fragment:02427" 5795..7987 /note="assembly_fragment:02438" 8088..135903 /note="assembly_fragment:02459" 136004..139240 /note="assembly_fragment:02490" 139241..142558 /note="assembly_fragment:02997" 33414 a 39141 c 33825 g 33673 t 505 others ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 6 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 2667: contig of 2667 bp in length * 2668 2767: gap of 100 bp * 2768 5694: contig of 2927 bp in length * 5695 5794: gap of 100 bp * 5795 7987: contig of 2193 bp in length * 7988 8087: gap of 100 bp * 8088 135903: contig of 127816 bp in length * 135904 136003: gap of 100 bp * 136004 139240: contig of 3237 bp in length * 139241 139341: gap of 100 bp * 139341 142558: contig of 3216 bp in length. Location/Qualifiers 1..142558 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /map="q42.11-q42.3" /clone="RP5-881p19" /clone_id="RPC1-5" 1..2667 /note="assembly_fragment:02282" 2768..5694 /note="assembly_fragment:02427" 5795..7987 /note="assembly_fragment:02438" 8088..135903 /note="assembly_fragment:02459" 136004..139240 /note="assembly_fragment:02490" 139241..142558 /note="assembly_fragment:02997" 33414 a 39141 c 33825 g 33673 t 505 others ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 6 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 2667: contig of 2667 bp in length * 2668 2767: gap of 100 bp * 2768 5694: contig of 2927 bp in length * 5695 5794: gap of 100 bp * 5795 7987: contig of 2193 bp in length * 7988 8087: gap of 100 bp * 8088 135903: contig of 127816 bp in length * 135904 136003: gap of 100 bp * 136004 139240: contig of 3237 bp in length * 139241 139341: gap of 100 bp * 139341 142558: contig of 3216 bp in length. Location/Qualifiers 1..142558 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /map="q42.11-q42.3" /clone="RP5-881p19" /clone_id="RPC1-5" 1..2667 /note="assembly_fragment:02282" 2768..5694 /note="assembly_fragment:02427" 5795..7987 /note="assembly_fragment:02438" 8088..135903 /note="assembly_fragment:02459" 136004..139240 /note="assembly_fragment:02490" 139241..142558 /note="assembly_fragment:02997" 33414 a 39141 c 33825 g 33673 t 505 others ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 6 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 2667: contig of 2667 bp in length * 2668 2767: gap of 100 bp * 2768 5694: contig of 2927 bp in length * 5695 5794: gap of 100 bp * 5795 7987: contig of 2193 bp in length * 7988 8087: gap of 100 bp * 8088 135903: contig of 127816 bp in length * 135904 136003: gap of 100 bp * 136004 139240: contig of 3237 bp in length * 139241 139341: gap of 100 bp * 139341 142558: contig of 3216 bp in length. Location/Qualifiers 1..142558 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /map="q42.11-q42.3" /clone="RP5-881p19" /clone_id="RPC1-5" 1..2667 /note="assembly_fragment:02282" 2768..5694 /note="assembly_fragment:02427" 5795..7987 /note="assembly_fragment:02438" 8088..135903 /note="assembly_fragment:02459" 136004..139240 /note="assembly_fragment:02490" 139241..142558 /note="assembly_fragment:02997" 33414 a 39141 c 33825 g 33673 t 505 others ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 6 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 2667: contig of 2667 bp in length * 2668 2767: gap of 100 bp * 2768 5694: contig of 2927 bp in length * 5695 5794: gap of 100 bp * 5795 7987: contig of 2193 bp in length * 7988 8087: gap of 100 bp * 8088 135903: contig of 127816 bp in length * 135904 136003: gap of 100 bp * 136004 139240: contig of 3237 bp in length * 139241 139341: gap of 100 bp * 139341 142558: contig of 3216 bp in length. Location/Qualifiers </p>					

alignment_scores:
Quality: 54.00 Length: 16
Ratio: 4.154 Gaps: 0
Percent Similarity: 81.250 Percent Identity: 50.000

alignment_block:

US-09-471-276-831_COPY_1_16 x AL136379/rev ..

Align seg 1/1 to reverse of: AL136379 from: 1 to: 142558

1 MetSetMetLeuValValpHeLeuLeuTPrpGlyValThrTrpGly 16

56577 GGTCTCATGATGATGCTTGTGTCTGTGAGCATCATCTGGCGGA 56530

seq_name: gb_hlcg:AC021458

seq_documentation_block:

LOCUS AC021458 152922 bp DNA HTG 16-MAR-2000
DEFINITION Homo sapiens clone RP11-222B11, WORKING DRAFT SEQUENCE, 15
unordered pieces.

ACCESSION AC021458

VERSION AC021458.3 GI:7249151

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE 1 (bases 1 to 152922)

JOURNAL Homo sapiens, clone RP11-222B11

REFERENCE 2 (bases 1 to 152922)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Batow, J., Barna, N., Beckert, R., Beda, F.,

Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,

Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,

DeRellano, K., Dewar, K., Domino, M., Doyle, M., Fenster, J.,

Garfield, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,

Howland, J. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J.,

Landers, T., Lehotzky, J., Levine, R., Liu, C., Liu, G., Locke, K.,

Macdonald, P., Marquis, N., McEwan, P., McQuirk, A., McKernan, K.,

McNeeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,

Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,

Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,

Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,

Stefanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,

Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,

Zimmer, A. and Zody, M.

Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 16, 2000 this sequence version replaced gi:6721264.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L5191

Center clone name: 222_B_11

Sequencing vector: M13; M77815: 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 144573 bases at least Q40

Consensus quality: 148358 bases at least Q30

Consensus quality: 149844 bases at least Q20

Insert size: 151000; agarose-ff

Insert size: 151522; sum-of-contigs

Quality coverage: 4.8 in Q20 bases; agarose-ff
Quality coverage: 4.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1022: contig of 1022 bp in length
* 1023 1122: gap of 100 bp
* 1123 2651: contig of 1529 bp in length
* 2652 2751: gap of 100 bp
* 2752 5441: contig of 2690 bp in length
* 5442 5541: gap of 100 bp
* 5542 7626: contig of 2085 bp in length
* 7627 7726: gap of 100 bp
* 7727 12283: contig of 4557 bp in length
* 12284 12383: gap of 100 bp
* 12384 16229: contig of 3846 bp in length
* 16230 16329: gap of 100 bp
* 16330 20224: contig of 3895 bp in length
* 20225 20324: gap of 100 bp
* 20325 27472: contig of 7148 bp in length
* 27473 27572: gap of 100 bp
* 27573 34532: contig of 6960 bp in length
* 34533 34632: gap of 100 bp
* 34633 44164: contig of 9532 bp in length
* 44165 44264: gap of 100 bp
* 44265 65244: contig of 20980 bp in length
* 65245 65344: gap of 100 bp
* 65345 83734: contig of 18390 bp in length
* 83735 83834: gap of 100 bp
* 83835 103507: contig of 19673 bp in length
* 103508 103607: gap of 100 bp
* 103608 129866: contig of 26259 bp in length
* 129867 129966: gap of 100 bp
* 129967 152922: contig of 22956 bp in length.

FEATURES

source

misc_feature
1..1022
/note="assembly_fragment"
misc_feature
1123..2651
/note="assembly_fragment"
misc_feature
2752..5441
/note="assembly_fragment"
misc_feature
5442..7626
/note="assembly_fragment"
misc_feature
7727..12283
/note="assembly_fragment"
misc_feature
12384..16229
/note="assembly_fragment"
misc_feature
16330..20224
/note="assembly_fragment"
misc_feature
20325..27472
/note="assembly_fragment"
misc_feature
27573..34532
/note="assembly_fragment"
misc_feature
34633..44164
/note="assembly_fragment"
misc_feature
44265..65244
/note="assembly_fragment"
misc_feature
65345..83734
/note="assembly_fragment"
misc_feature
83835..103507
/note="assembly_fragment"

```

/note "assembly_fragment"
103608 .129866
/note "assembly_fragment"
clone_end:SP6
vector_side:right"
misc_feature
129967 .152922
/note "assembly_fragment"
BASE COUNT 49092 a 29431 c 28272 g 44723 t 1404 others
ORIGIN

alignment_scores:
Quality: 54.00 Length: 15
Ratio: 4.500 Gaps: 0
Percent Similarity: 80.000 Percent Identity: 53.333

alignment_block:
US-09-471-276-831_COPY_1_16 x AC021458/rev ..
Align seg 1/1 to reverse of: AC021458 from: 1 to: 152922

2 SerMetLeuValAlpheLeuLeuTrpGlyValThrTrpGly 16
||||: |||||: |||||: |||||: |||||
15271 AGTATATAAAATGCTTTTATCATCATCTGGGACCTGCTGGGCT 15227

seq_name: gb_htg:AC092729

seq_documentation_block:
LOCUS AC092729 163977 bp DNA HTG 21-JUL-2001
DEFINITION Canis familiaris clone RP81-60B6, WORKING DRAFT SEQUENCE, 8
unorderd pieces.
AC092729
AC092729.1 GI:14993720
HTG: HTGS_PHASE1: HTGS_DRAFT.
SOURCE
dog
ORGANISM
Canis familiaris
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 163977)
Avele, R., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Boulford, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L.,
Granite, S., Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E.,
Lee-Lin, S.-O., Legaspi, R., Lim, M., Maduro, O.L., Maduro, V.B.,
Masigallo, C., Mastrian, S.D., McCloskey, J.C., McDowell, J.,
Pearson, R., Prasad, A., Shevchenko, Y., Snyder, B., Stantrop, S.,
Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgou, C., Vogt, J.L.,
Walker, M.A., Wetherby, K.D., Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 163977)
Green, E.D.
Direct Submission
Submitted (21-JUL-2001) NIH Intramural Sequencing Center, 8717
GroveMont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@hgrl.nih.gov
----- Project Information
Center project name: c1g
Center clone name: 060B06
----- Summary Statistics
Sequencing vector: plasmid: n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160809 bases at least Q40
Consensus quality: 161490 bases at least Q30
Consensus quality: 161950 bases at least Q20
Insert size: 152000; agarose-fp
Insert size: 163277; sum-of-contigs
Quality coverage: 11.00x in Q20 bases; agarose-fp
Quality coverage: 10.24x in Q20 bases; sum-of-contigs

```

```

*****
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*
* 1 3104: contig of 3104 bp in length
* 3105 gap of unknown length
* 3205 7098: contig of 3694 bp in length
* 7099 gap of unknown length
* 7199 14995: contig of 7797 bp in length
* 14996 15095: gap of unknown length
* 25323 25423: contig of 10228 bp in length
* 25324 25423: gap of unknown length
* 25424 43145: contig of 17722 bp in length
* 43146 43245: gap of unknown length
* 43246 71672: contig of 28427 bp in length
* 71673 71772: gap of unknown length
* 71773 108480: contig of 36708 bp in length
* 108481 108580: gap of unknown length
* 108581 163977: contig of 55397 bp in length.
Location/Qualifiers
1..163977
/organism:"Canis familiaris"
/db_xref:"taxon:9615"
/clone:"RP81-60B6"
/clone_1lb:"RP81"
1..3104
/note:"assembly_fragment"
3205..7098
/note:"assembly_fragment"
clone_end:77
vector_side:left"
7199..14995
/note:"assembly_fragment"
15096..25323
/note:"assembly_fragment"
25424..43145
/note:"assembly_fragment"
clone_end:SP6
vector_side:right"
43246..71672
/note:"assembly_fragment"
71773..108480
/note:"assembly_fragment"
108581..163977
/note:"assembly_fragment"
BASE COUNT 39995 a 40846 c 41088 g 41340 t 708 others
ORIGIN

alignment_scores:
Quality: 54.00 Length: 13
Ratio: 4.909 Gaps: 0
Percent Similarity: 84.615 Percent Identity: 69.231

alignment_block:
US-09-471-276-831_COPY_1_16 x AC092729 ..
Align seg 1/1 to: AC092729 from: 1 to: 163977

4 LeuValValAlpheLeuLeuTrpGlyValThrTrpGly 16
||||: |||||: |||||: |||||: |||||
31963 TTGACACTGACAAGACTATTATGCGCAATTACATGGCGC 32001

seq_name: gb_pr:AL357115

seq_documentation_block:
LOCUS AL357115 164012 bp DNA PRI 02-FEB-2001
DEFINITION Human DNA sequence from clone RP11-102P23 on chromosome

```


Xq13.3-21.1, complete sequence.

ACCESSION AL512502.4
 VERSION AL512502.4
 KEYWORDS HTGS:12666259
 SOURCE HTG.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 164012)

REFERENCE 1
 AUTHORS Wray, P.
 TITLE Direct Submission
 JOURNAL Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Feb 5, 2001 this sequence version replaced gi:12323354.

COMMENT During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX> RP11-102P23 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6

FEATURES
 source This sequence is the entire insert of clone RP11-102P23 The true left end of clone RP11-346E8 is at 115722 in this sequence.
 Location/Qualifiers
 1..164012
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="q13.3-21.1"
 /clone="RP11-102P23"
 /clone_lib="RPCI-11.1"
 BASE COUNT 48623 a 29015 c 31474 g 54900 t
 ORIGIN

alignment_scores:
 Quality: 54.00 Length: 11
 Ratio: 5.400 Gaps: 0
 Percent Similarity: 90.909 Percent Identity: 72.727

alignment_block:
 US-09-471-276-831_COPY_1_16 x AL512502 ..
 Align seg 1/1 to: AL512502 from: 1 to: 164012

6 ValPheLeuLeuLeuTrpGlyValThrTrpGly 16
 |||:::|||||||:::|||||
 3282 GTTACCTTTACTTGGCGTATTGCTGCGCG 3314

seq_name: gp_htg:AL512502

seq_documentation_block:
 LOCUS AL512502 177263 bp DNA HTG 10-JUL-2001
 DEFINITION Homo sapiens chromosome X clone RP11-807E6, *** SEQUENCING IN
 PROGRESS ***; 6 unordered pieces.

ACCESSION AL512502
 VERSION AL512502.4
 KEYWORDS HTGS:14456353
 SOURCE HTGS_PHASE1: HTGS_CANCELLED.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 177263)

REFERENCE 1
 AUTHORS Heath, P.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jun 15, 2001 this sequence version replaced gi:13092346.

COMMENT ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BA807E6
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid: 108752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 175699 bases at least Q40
 Consensus quality: 176106 bases at least Q30
 Consensus quality: 176347 bases at least Q20
 Insert size: 176763; sum-of-contigs
 Insert size: 174694; 7.7% error; agarose-fp
 Quality coverage: 7.03x in Q20 bases; sum-of-contigs Quality
 coverage: 7.16x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1
 * 8339: contig of 8339 bp in length
 * 8340 8439: gap of 100 bp
 * 8440 69040: contig of 60601 bp in length
 * 69041 69140: gap of 100 bp
 * 69141 108593: contig of 39453 bp in length
 * 108594 108693: gap of 100 bp
 * 108694 113120: contig of 4427 bp in length
 * 113121 113220: gap of 100 bp
 * 113221 148811: contig of 35591 bp in length
 * 148812 148911: gap of 100 bp
 * 148912 177263: contig of 28352 bp in length.
 Location/Qualifiers
 1..177263
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /clone="RP11-807E6"
 /clone_lib="RPCI-11.3"
 1..8339
 /note="assembly_fragment:01802
 fragment_chain:1"
 8440..69040
 /note="assembly_fragment:02140
 fragment_chain:1"
 69141..108593
 /note="assembly_fragment:03321
 fragment_chain:1"
 108694..113120
 /note="assembly_fragment:00292
 fragment_chain:1"
 113221..148811
 /note="assembly_fragment:01412

FEATURES
 source

misc_feature
 1..8339
 /note="assembly_fragment:01802
 fragment_chain:1"
 8440..69040
 /note="assembly_fragment:02140
 fragment_chain:1"
 69141..108593
 /note="assembly_fragment:03321
 fragment_chain:1"
 108694..113120
 /note="assembly_fragment:00292
 fragment_chain:1"
 113221..148811
 /note="assembly_fragment:01412

misc_feature
 108694..113120
 /note="assembly_fragment:00292
 fragment_chain:1"
 113221..148811
 /note="assembly_fragment:01412

COMMENT

On Jun 20, 2000 this sequence version replaced g1:8573783. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RP3-380B8 is from the library RP3-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2> This sequence is the entire insert of clone RP3-380B8. The true left end of clone RP1-525021 is at 141303 in this sequence. The true right end of clone RP1-182016 is at 57138 in this sequence.

FEATURES

source

1.186882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p24.1-25.3"
/clone="RP3-380B8"
/clone_lib="RP3-3"
247..465
/note="MIR repeat: matches 16..236 of consensus"
MIR repeat: matches 16..236 of consensus"
1442..2391
/note="LIMB3 repeat: matches 5281..6174 of consensus"
LIMB3 repeat: matches 5281..6174 of consensus"
2435..2499
/note="MIR repeat: matches 77..145 of consensus"
MIR repeat: matches 77..145 of consensus"
2540..3786
/note="L1PA2 repeat: matches 4900..6146 of consensus"
2542..3786
/note="L1PA5 repeat: matches 4900..6143 of consensus"
4034..4653
/note="L1MA9 repeat: matches 4022..4632 of consensus"
L1MA9 repeat: matches 4022..4632 of consensus"
4654..4953
/note="AluX repeat: matches 1..300 of consensus"
AluX repeat: matches 1..300 of consensus"
4954..6496
/note="L1MA9 repeat: matches 4632..6308 of consensus"
L1MA9 repeat: matches 4632..6308 of consensus"
6522..6820
/note="MUTID repeat: matches 14..283 of consensus"
MUTID repeat: matches 14..283 of consensus"
6821..6990
/note="MR67A repeat: matches 374..543 of consensus"
MR67A repeat: matches 374..543 of consensus"
6987..7184
/note="MUTID repeat: matches 308..505 of consensus"
MUTID repeat: matches 308..505 of consensus"
7290..7462
/note="MIR repeat: matches 71..262 of consensus"
MIR repeat: matches 71..262 of consensus"
8007..8093
/note="MIR repeat: matches 71..177 of consensus"

/note="L1M4 repeat: matches 5688..5772 of consensus"
L1M4 repeat: matches 5688..5772 of consensus"
8094..8229
/note="L1MB8 repeat: matches 6029..6173 of consensus"
L1MB8 repeat: matches 6029..6173 of consensus"
8932..8994
/note="L2 repeat: matches 2637..2699 of consensus"
8996..9199
/note="MER33 repeat: matches -9..192 of consensus"
MER33 repeat: matches -9..192 of consensus"
9200..9559
/note="THE1B repeat: matches 1..364 of consensus"
THE1C repeat: matches 1..371 of consensus"
9560..9685
/note="MER33 repeat: matches 192..324 of consensus"
MER33 repeat: matches 192..324 of consensus"
10364..10771
/note="MUTIG repeat: matches 65..512 of consensus"
MUTIG repeat: matches 65..512 of consensus"
10775..10875
/note="MIR repeat: matches 134..252 of consensus"
12145..12963
/note="L1MA10 repeat: matches 5479..6315 of consensus"
L1MA10 repeat: matches 5479..6315 of consensus"
12979..13263
/note="AluX repeat: matches 9..280 of consensus"
AluX repeat: matches 9..290 of consensus"
13548..13778
/note="MIR repeat: matches 15..262 of consensus"
MIR repeat: matches 15..262 of consensus"
14035..14188
/note="MER5A repeat: matches 43..189 of consensus"
MER5A repeat: matches 43..189 of consensus"
14785..14859
/note="L1ME3A repeat: matches 5785..5859 of consensus"
L1ME repeat: matches 5785..5859 of consensus"
15050..15303
/note="L1ME3A repeat: matches 5525..5777 of consensus"
L1ME repeat: matches 5525..5777 of consensus"
15317..15774
/note="L1M4 repeat: matches 3951..4454 of consensus"
15337..15774
/note="L1 repeat: matches 3951..4433 of consensus"
15943..16022
/note="L2 repeat: matches 2654..2750 of consensus"
/note="match: GSS: Em:AQ541664"
17138..17193
/note="MIR repeat: matches 78..133 of consensus"
MIR repeat: matches 78..133 of consensus"
18611..18904
/note="MUTIC repeat: matches 193..466 of consensus"
MUTIC repeat: matches 193..466 of consensus"
18905..19230
/note="AluSg repeat: matches 1..310 of consensus"
AluSg repeat: matches 1..310 of consensus"
19231..19419
/note="MUTIC repeat: matches 1..193 of consensus"
MUTIC repeat: matches 1..193 of consensus"
join(<19542..19799,20379..>20477)
/gene="dJ380B8.3"
/note="match: ESTs: Em:R79825"
/evidence-not-experimental
/product="dJ380B8.3 (a novel mRNA)"
join(19542..19799,20379..20477)
19669..20180
/note="match: GSS: Em:AQ401011"
20155..20184
/note="15 copies 2 mer ac 100% conserved"
20482..20806
/note="L1R16C repeat: matches 56..387 of consensus"
L1R16C repeat: matches 56..387 of consensus"

repeatL_region	21088..21177 /note:"FLAM_A repeat: matches 10. .99 of consensus FLAM_A repeat: matches 10. .99 of consensus"
repeatL_region	21469..21856 /note:"HERVL repeat: matches 4644. .5045 of consensus" /note:"HERV.L repeat: matches 4644. .5045 of consensus"
repeatL_region	21469..22143 /note:"HERVL repeat: matches 4345. .5045 of consensus" /note:"HERV.L repeat: matches 4345. .5045 of consensus"
repeatL_region	22192..23403 /note:"HERVL repeat: matches 2953. .4189 of consensus HERVL repeat: matches 2992. .4228 of consensus"
repeatL_region	23429..24388 /note:"HERVL repeat: matches 1902. .2891 of consensus" /note:"HERV.L repeat: matches 1902. .2891 of consensus"
repeatL_region	23611..24388 /note:"HERVL repeat: matches 1902. .2701 of consensus" /note:"HERV.L repeat: matches 1902. .2701 of consensus"
repeatL_region	24949..25225 /note:"HERVL repeat: matches 882. .1164 of consensus HERVL repeat: matches 882. .1164 of consensus"
repeatL_region	25788..26094 /note:"Aluv repeat: matches 1. .307 of consensus Aluv repeat: matches 1. .307 of consensus"
repeatL_region	26354..26477 /note:"LIME repeat: matches 458. .585 of consensus" /note:"L2 repeat: matches 2666. .2741 of consensus"
repeatL_region	26716..26792 /note:"L2 repeat: matches 2666. .2741 of consensus" complement(26932. .27550) /note:"match: GSS: Em:A0417969"
misc_feature	27047..27174 /note:"MIR repeat: matches 65. .196 of consensus MIR repeat: matches 65. .196 of consensus"
repeatL_region	27177..27238 /note:"L2 repeat: matches 2644. .2709 of consensus L2 repeat: matches 2684. .2749 of consensus"
repeatL_region	27243..27405 /note:"L2 repeat: matches 2576. .2750 of consensus"
alignment_scores:	Quality: 54.00 Length: 13 Ratio: 4.909 Gaps: 0 Percent Similarity: 84.615 Percent Identity: 61.538
alignment_block:	US-09-471-276-831-COPY_1_16 x ALJ36307/rev .. Align seg 1/1 to reverse of: ALJ36307 from: 1 to: 186882
seq_name: gb_hltg:AC012285	4 LeuValValPheLeuLeuLeuTPrGlyValThrTrpGly 16 : :::
seq_documentation_block:	LOCUS AC012285 283388 bp DNA HTG 22-OCT-1999 DEFINITION Homo sapiens clone 3_F_23.*** SEQUENCING IN PROGRESS ***, 34 unordered pieces. ACCESSION AC012285 VERSION AC012285.1 GI:6094585 KEYWORDS HTG: HTGS_PHASE1. SOURCE human. ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 283388)
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Homo sapiens, clone 3_F_23
JOURNAL	Unpublished
REPEAT_REGIONS	2 (bases 1 to 283388) Baldwin,J., Barra,N., Beckerly,R., Boukslavsky,L., Bouhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collimore,A., Cook,P., Dearellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
COMMENT	Galgan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,T., Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Lien,C., Locke,K., Macdonald,P., Margus,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Mejdrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talams,J., Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M. Direct Submission Submitted (22-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html. NOTE: This is a 'working draft' sequence. It currently consists of 54 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1	2094: contig of 2094 bp in length
2095	gap of unknown length
4250	contig of 2155 bp in length
6745	gap of unknown length
9598	contig of 2495 bp in length
12052	gap of unknown length
14199	contig of 2853 bp in length
16646	gap of unknown length
18763	contig of 2454 bp in length
21642	gap of unknown length
23923	contig of 2147 bp in length
26405	gap of unknown length
29548	contig of 2447 bp in length
31971	gap of unknown length
35590	contig of 2117 bp in length
38764	gap of unknown length
42512	contig of 2482 bp in length
45512	gap of unknown length
50111	contig of 2174 bp in length
55793	gap of unknown length
58925	contig of 2281 bp in length
61301	gap of unknown length
64378	contig of 2482 bp in length
69332	gap of unknown length
71903	contig of 2174 bp in length
74966	gap of unknown length

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (11-Jul-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2001 this sequence version replaced gi:12656690.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----

Contig length: 111131
Phrap values in estimate: 110454
Average error rate (BCM-Phrap estimate): 1.53449e-05
Fraction of Phrap values less than 40 : 0.0102336
Number of N's in consensus : 52
0

----- Consensus changing edits -----

Position	Original+Context	Edited+Context
5038	ctgagatgc(n)ccatgcacat	ctgagatgc(a)ccatgcacat
11091	ctattagacc(n)tagaccntg	ctattagacc(t)tagaccntg
11098	gccttagacc(n)tggtcgtgt	gccttagacc(t)tggtcgtgt
11107	ctgtgctcg(n)ttttcttct	ctgtgctcg(t)ttttcttct
17387	ccctggcaac(n)cgttgaaac	ccctggcaac(a)cgttgaaac
19617	tgctctctaa(n)tttttaaat	tgctctctaa(t)tttttaaat
20451	ctgtgtgaac(n)atagttga	ctgtgtgaac(a)atagttga
21493	ttctcact(n)tttattagct	ttctcact(c)tttattagct
24247	gggcagggca(n)ggcagggcag	gggcagggca(g)ggcagggcag
24289	gaagggcag(n)canngcaggg	gaagggcag(g)canngcaggg
24292	gggcagggca(n)ggcagggcag	gggcagggca(g)ggcagggcag
34999	tttttttt(n)aaagagagat	tttttttt(t)aaagagagat
35002	tttttttaa(n)aaagagatc	tttttttaa(a)aaagagatc
35004	tttttaana(n)ggagcttcac	tttttaana(a)ggagcttcac
35039	tgagttgcag(n)gggtctatct	tgagttgcag(g)gggtctatct
35126	caactcacag(n)gcacacacac	caactcacag(t)gcacacacac
35164	gtatttttag(n)agagatgggg	gtatttttag(t)agagatgggg

47108	atataccctaa(n)ngaaagaaaa	atataccctaa(a)ngaaagaaaa
47109	tatacctaan(n)gaaagaaat	tatacctaa(a)gaaagaaat
47731	ggaaaaaaca(n)tgtaaat	ggaaaaaaca(a)tgtaaat
56868	caccacacat(n)aaccacacag	caccacacat(t)aaccacacag
56893	tlaagatctc(n)tttaggtlan	tlaagatctc(t)tttaggtlan
56894	taagatctcn(n)tttaggtlan	taagatctc(t)tttaggtlan
56899	tcctnmttag(n)gttagtttag	tcctnmttag(t)gttagtttag
56903	nttaggtta(n)gttagtttag	nttaggtta(t)gttagtttag
56916	ttattcag(n)tcgngcct	ttattcag(t)tcgngcct
56921	tcagttcag(n)gcanltat	tcagttcag(t)gcanltat
56925	tttatctc(n)ttatattla	tttatctc(t)ttatattla
56941	tttatctc(n)caatataat	tttatctc(t)caatataat
56942	caagagatag(n)caatataat	caagagatag(a)caatataat
56998	gtatttgga(n)tatatttac	gtatttgga(c)tatatttac
57135	cactcattt(n)taagtgat	cactcattt(c)taagtgat
57224	caaaaataa(n)taaacacac	caaaaataa(g)taaacacac
57233	antaaaac(n)caatttccc	antaaaac(a)caatttccc
60057	tggttgaat(n)tttacctct	tggttgaat(t)tttacctct
69183	tgaaacagta(n)nnnctaga	tgaaacagta(a)nnnctaga
69184	gaacagctaa(n)nnnctaga	gaacagctaa(a)nnnctaga
69185	aacagctann(n)nnctaga	aacagctann(a)nnnctaga
69186	acagctann(n)nnctaga	acagctann(a)nnnctaga
69187	acagctann(n)nnctaga	acagctann(a)nnnctaga
69188	agctannnn(n)ctagaaggg	agctannnn(a)ctagaaggg
72601	agcttgcctt(n)atctgactaa	agcttgcctt(a)atctgactaa
73530	atcaaacgtg(n)ttcccttga	atcaaacgtg(t)ttcccttga
79249	aaacagagat(n)tagatcaat	aaacagagat(t)tagatcaat
79276	aaacagagcc(n)tagatcaat	aaacagagcc(t)tagatcaat
79277	acagagcccn(n)agaaataac	acagagcccn(a)agaaataac
79299	tgcatatcta(n)actatctga	tgcatatcta(c)actatctga
79300	gcatatctaa(n)actatctga	gcatatctaa(a)actatctga
79310	nactatctga(n)cttgcacaa	nactatctga(t)cttgcacaa
79316	ctganctttg(n)caaacctgag	ctganctttg(a)caaacctgag
79337	aaacacagc(n)atggggaag	aaacacagc(a)atggggaag
86864	cccaaggag(n)tgataagctgg	cccaaggag(c)tgataagctgg

----- Distribution of Quality < 40 Bases -----

#	bases	5	10	15	20	25	30	35	40
1000	1000								
900	900								
800	800								
700	700								
600	600								
500	500								
400	400								
300	300								
200	200								
100	100								
0	0	*	*	*	*	*	*	*	*

Phrap Value Range

----- Version: 1.01 gxi0. -----

FEATURES Location/Qualifiers
source 1..111131

repeat_region
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-847A19"
complement(516..741)
/rpt_family="AluDo"

alignment_scores:
Quality: 53.00 Length: 15
Ratio: 4.417 Gaps: 0
Percent Similarity: 80.000 Percent Identity: 53.333

alignment_block:
us-09-471-276-831_copy_1_16 x AC046131/rev ..


```

Center project name: L8917
Center clone name: 21.G.15
----- Summary Statistics -----
Sequencing vector: M13; M77815: 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 13367 bases at least Q40
Consensus quality: 140370 bases at least Q30
Consensus quality: 142846 bases at least Q20
Insert size: 193000; agarose-fp
Insert size: 144056; sum-of-ctrls
Quality coverage: 3.1 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 15351: contig of 15351 bp in length
* 15352 15451: gap of 100 bp
* 15452 16828: contig of 1377 bp in length
* 16829 16928: gap of 100 bp
* 16929 17939: contig of 1011 bp in length
* 17940 18039: gap of 100 bp
* 18040 19688: contig of 1649 bp in length
* 19689 19788: gap of 100 bp
* 19789 21805: contig of 2017 bp in length
* 21806 21905: gap of 100 bp
* 21906 23993: contig of 2088 bp in length
* 23994 24093: gap of 100 bp
* 24094 27304: contig of 3211 bp in length
* 27305 27404: gap of 100 bp
* 27405 32590: contig of 5186 bp in length
* 32591 32690: gap of 100 bp
* 32691 37217: contig of 4527 bp in length
* 37218 37317: gap of 100 bp
* 37318 43108: contig of 5791 bp in length
* 43109 43208: gap of 100 bp
* 43209 47763: contig of 4555 bp in length
* 47764 47863: gap of 100 bp
* 47864 53820: contig of 5957 bp in length
* 53821 53920: gap of 100 bp
* 53921 58914: contig of 4994 bp in length
* 58915 59014: gap of 100 bp
* 59015 85355: contig of 26341 bp in length
* 85356 85455: gap of 100 bp
* 85456 94988: contig of 9533 bp in length
* 94989 95088: gap of 100 bp
* 95089 109604: contig of 14516 bp in length
* 109605 109704: gap of 100 bp
* 109705 122428: contig of 12724 bp in length
* 122429 122528: gap of 100 bp
* 122529 139527: contig of 16999 bp in length
* 139528 139627: gap of 100 bp
* 139628 145856: contig of 6229 bp in length.
Location/Qualifiers
1. 145856
/organism: "Homo sapiens"
/db_xref: "taxon:9606"
/chromosome: "18"
/map: "18"
/clone: "RP11-21G15"
/clone_lib: "RP11-11 Human Male BAC"
1. 15351
/note: "assembly-fragment"
clone_end: SP6
vector_side: left"
misc_feature
15452. .16828
/note: "assembly-fragment"

```

```

misc_feature 16929. .17939
/note: "assembly-fragment"
misc_feature 18040. .19688
/note: "assembly-fragment"
misc_feature 19789. .21805
/note: "assembly-fragment"
misc_feature 21906. .23993
/note: "assembly-fragment"
misc_feature 24094. .27304
/note: "assembly-fragment"
misc_feature 27405. .32590
/note: "assembly-fragment"
misc_feature 32691. .37217
/note: "assembly-fragment"
misc_feature 37318. .43108
/note: "assembly-fragment"
misc_feature 43209. .47763
/note: "assembly-fragment"
misc_feature 47864. .53820
/note: "assembly-fragment"
misc_feature 53921. .58914
/note: "assembly-fragment"
misc_feature 59015. .85355
/note: "assembly-fragment"
misc_feature 85456. .94988
/note: "assembly-fragment"
misc_feature 95089. .109604
/note: "assembly-fragment"
misc_feature 109705. .122428
/note: "assembly-fragment"
misc_feature 122529. .139527
/note: "assembly-fragment"
misc_feature 139628. .145856
/note: "assembly-fragment"
clone_end: 17
vector_side: right"

BASE COUNT 41606 a 30265 c 28947 g 43233 t 1805 others
ORIGIN
alignment_scores:
Quality: 53.00 Length: 15
Ratio: 4.818 Gaps: 0
Percent Similarity: 73.333 Percent Identity: 60.000

alignment block:
US-09-471-276-831_COPY_1_16 x AC073948 ..
Align seg 1/1 to: AC073948 from: 1 to: 145856
2 SerMetLeuValAlaPheLeuLeuTrpGlyValIhrTrpGly 16
||||||| ::||| ||||| |||||
14470 TCAGTCTGTATATCTTCTTCTTAGTTGAGAACGACGAGGGGT 14514

seq_name: gb_hlg:AC019241
seq_documentation block:
LOCUS AC019241 155804 bp DNA HTG 10-NOV-2000
DEFINITION Homo sapiens clone RP11-45120, WORKING DRAFT SEQUENCE, 6 unordered
pieces.
ACCESSION AC019241
VERSION AC019241.5 GI:11136802
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 155804)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-45120
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155804)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

```



```

Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G., Castle,A.,
Chopel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Deatellano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J.,
Ferrelta,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marcuis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Menus,L., Morrow,J., Naylor,D.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A., and Zody,M.

TITLE
JOURNAL
COMMENT
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1167
Center clone name: 45_I-20
----- Summary Statistics
Sequencing vector: M13: M77815; 47% of reads
Sequencing vector: Plasmid; n/a; 53% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 154252 bases at least Q40
Consensus quality: 154676 bases at least Q30
Consensus quality: 155008 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 155304; sum-of-contents
Quality coverage: 11.1 in Q20 bases; agarose-fp
Quality coverage: 11.7 in Q20.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 41619: contig of 41619 bp in length
* 41620 41719: gap of 100 bp
* 41720 47544: contig of 5825 bp in length
* 47545 47644: gap of 100 bp
* 47645 59936: contig of 12292 bp in length
* 59937 60036: gap of 100 bp
* 60037 83161: contig of 23145 bp in length
* 83162 83281: gap of 100 bp
* 83282 117655: contig of 34374 bp in length
* 117656 117755: gap of 100 bp
* 117756 155804: contig of 38049 bp in length.

FEATURES
source
1..155804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-45120"
/clone_lib="RPCT-11 Human Male PAC"
1..41619
/misc_feature
/Note="assembly-fragment"
clone_end:5825
vector_side:left"
misc_feature
41720..47544
/Note="assembly-fragment"
47645..59936
/misc_feature
60037..83161
/misc_feature
83282..117655
/misc_feature
117756..155804
/Note="assembly-fragment"
clone_end:17
vector_side:right"
500 others

BASE COUNT 46801 a 31756 c 31189 g 45558 t
ORIGIN

alignment_scores:
Quality: 53.00 Length: 13
Ratio: 4.818 Gaps: 0
Percent Similarity: 84.615 Percent Identity: 61.538

alignment_block:
US-09-471-276-831_COPY_1_16 x AC019241/rev ..
Align seg 1/1 to reverse of: AC019241 from: 1 to: 155804
3 MetLeuValAlaPheLeuLeuTrpGlyValIhrTrp 15
||||:||||| |||:||||| |||:||||| |||:|||||
35549 ATGATGCTGTGCTTTGTCTTTGTGGGACTGAGGTGG 35511
seq_name: gb_hhg:AC025621

seq_documentation_block:
LOCUS AC025621.160294 bp DNA HTG 14-DEC-2000
DEFINITION Homo sapiens clone RP11-29E15, WORKING DRAFT SEQUENCE, 8 unordered
pieces.
ACCESSION AC025621
VERSION AC025621.3 GI:11761489
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 160294)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL
TITLE Homo sapiens, clone RP11-29E15
UNPUBLISHED
2 (bases 1 to 160294)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,A., Beda,F.,
Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Chopel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferrelta,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-pierre,C., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marcuis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Menus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggili,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 14, 2000 this sequence version replaced gi:11181821.
All repeats were identified using RepeatMasker:

```

Smtl, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 29.E.15
 Center clone name: 29.E.15

----- Summary Statistics
 Sequencing vector: M13: M77815; % of reads
 Sequencing vector: Plasmid: n/a; 97% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 157651 bases at least Q40
 Consensus quality: 158996 bases at least Q30
 Consensus quality: 159393 bases at least Q20
 Insert size: 163000; agarose-fp
 Insert size: 159594; sum-of-ctrls
 Quality coverage: 7.3 in Q20 bases; agarose-fp
 Quality coverage: 7.4 in Q20 bases.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1      121: contig of 121 bp in length
*      122 221: gap of 100 bp
*      222      2176: contig of 1955 bp in length
*      2177 2276: gap of 100 bp
*      2277 5789: contig of 3513 bp in length
*      5790 5889: gap of 100 bp
*      5890 11465: contig of 5576 bp in length
*      11466 11565: gap of 100 bp
*      11566 22920: contig of 11355 bp in length
*      22921 23020: gap of 100 bp
*      23021 43085: contig of 20065 bp in length
*      43086 43185: gap of 100 bp
*      43186 85707: contig of 42522 bp in length
*      85708 85807: gap of 100 bp
*      85808 160294: contig of 74487 bp in length.
  
```

FEATURES
 source
 1..160294
 /organism:"Homo sapiens"
 /db_xref="taxon:9606"
 /clone:"RP11-29E15"
 /clone_1lb:"RPct-11 Human Male BAC"
 1..121
 /note:"assembly-fragment"
 clone_end:SP6
 vector_side:left
 222..2176
 /note:"assembly-fragment"
 2277..5789
 /note:"assembly-fragment"
 5890..11465
 /note:"assembly-fragment"
 11566..22920
 /note:"assembly-fragment"
 23021..43085
 /note:"assembly-fragment"
 43186..85707
 /note:"assembly-fragment"
 85808..160294
 /note:"assembly-fragment"
 BASE COUNT 49637 a 29398 c 29387 g 51172 t 700 others

alignment_scores:
 Quality: 53.00 Length: 10
 Ratio: 5.300 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 70.000

alignment_block:
 US-09-471-276-831_COPY_1_16 x AC025621 ..
 Align seg 1/1 to: AC025621 from: 1 to: 160294

6 ValPheLeuLeuTrpGlyValThrTrp 15
 ::::|||||:|||||:|||||:|||||:|||||
 66325 CTTTTCCTAATTCGTGGGAATTCATGCG 66354

seq_name: gb_hcg:AC073872

seq_documentation_block:
 LOCUS AC073872 162080 bp DNA HTG 15-JUL-2000
 DEFINITION Homo sapiens chromosome 12 clone RP11-546D23, WORKING DRAFT
 SEQUENCE, 15 unordered pieces.
 AC073872
 AC073872.1 GI:8844170
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 162080)
 Waterston, R.H.
 The sequence of Homo sapiens clone
 2 (bases 1 to 162080)
 Waterston, R.H.
 Direct Submission
 Submitted (30-JUN-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

COMMENT
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 ----- Project Information -----
 Center project name: H.NH0546D23
 ----- Summary Statistics -----
 Sequencing vector: M13: 100%
 Sequencing vector: Plasmid: 0%
 Chemistry: Dye-primer ET; 100% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 152231 bases at least Q40
 Consensus quality: 155172 bases at least Q30
 Consensus quality: 156840 bases at least Q20
 Insert size: 162000; agarose-fp
 Insert size: 160680; sum-of-ctrls
 Quality coverage: 4.57 in Q20 bases; agarose-fp
 Quality coverage: 4.54 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1      1866: contig of 1866 bp in length
*      1867      1966: gap of unknown length
*      1967      3659: contig of 1693 bp in length
*      3660      3759: gap of unknown length
*      3760      5683: contig of 1924 bp in length
*      5684      8308: gap of unknown length
*      8308: contig of 2525 bp in length
  
```

```

*      8309      8408: gap of unknown length
*      8409      12809: contig of 4401 bp in length
*      12810      12909: gap of unknown length
*      12910      22137: contig of 9228 bp in length
*      22138      22237: gap of unknown length
*      22238      31506: contig of 9269 bp in length
*      31507      31606: gap of unknown length
*      31607      40674: contig of 9068 bp in length
*      40675      40774: gap of unknown length
*      40775      51622: contig of 10848 bp in length
*      51623      51722: gap of unknown length
*      51723      61563: contig of 9841 bp in length
*      61564      61663: gap of unknown length
*      61664      77409: contig of 15746 bp in length
*      77410      77509: gap of unknown length
*      77510      91451: contig of 13942 bp in length
*      91452      91551: gap of unknown length
*      91552      105478: contig of 13927 bp in length
*      105479      105578: gap of unknown length
*      105579      120820: contig of 15242 bp in length
*      120821      120920: gap of unknown length
*      120921      162080: contig of 41160 bp in length.
FEATURES
    source
        1. 162080
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="12"
            /clone="RP11-546D23"
            1. 1866
                /note="assembly_name:Contig3"
            1967. 3659
                /note="assembly_name:Contig4"
            3760. 5683
                /note="assembly_name:Contig5"
            5784. 8308
                /note="assembly_name:Contig6"
            8409. 12809
                /note="assembly_name:Contig7"
            12910. 22137
                /note="assembly_name:Contig8"
            22238. 31506
                /note="assembly_name:Contig9"
            31607. 40674
                /note="assembly_name:Contig10"
            clone_end:T7
            vector_side:left"
            40775. 51622
                /note="assembly_name:Contig11"
            clone_end:SP6
            vector_side:left"
            51723. 61563
                /note="assembly_name:Contig12"
            61664. 77409
                /note="assembly_name:Contig13"
            77510. 91451
                /note="assembly_name:Contig14"
            91552. 105478
                /note="assembly_name:Contig15"
            105579. 120820
                /note="assembly_name:Contig16"
            120921. 162080
                /note="assembly_name:Contig17"
BASE COUNT      53242 a 28804 c 27772 g 50759 t      1503 others
ORIGIN

```

```

alignment_scores:
    Quality:      53.00      Length:      15
    Ratio:        4.417      Gaps:      0
    Percent Similarity: 80.000      Percent Identity: 53.333
alignment_block:
    US-09-471-276-831_COPY_1_16 x AC073872      ..

```

```

Align seg 1/1 to: AC073872 from: 1 to: 162080
2 SerNetLeuValIAlpHeleuLeuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||:|||||
94379 TCCCTATATGTCGATTTACTGATCTGGGGTTTGTGGGG 94423
seq_name: em_htg_hum:AC011936
seq_documentation_block:
ID      AC011936      standard; DNA; HTG; 165689 BP.
XX
AC      AC011936;
XX
SV      AC011936.4
XX
DT      18-OCT-1999 (Rel. 61, Created)
DT      03-MAR-2000 (Rel. 62, Last updated, Version 4)
XX
DE      Homo sapiens clone RP11-16D24, WORKING DRAFT SEQUENCE, 19 unordered pieces.
XX
KW      HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS      Homo sapiens (human)
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC      Eutheria; Primates; Catarrhini; Hominiidae; Homo.
XX
[1]
RP      1-165689
RA      Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,
RA      Baldwin J., Barna N., Beckerly R., Boguslavsky L., Boukhgalter B.,
RA      Brown A., Castle A., Colangelo M., Collins S., Collymore A., Cooke P.,
RA      Dearlano K., Dewar K., Domino M., Donegan L., Doyle M., Ferreira P.,
RA      Fitzhugh W., Forrest C., Funke R., Gage D., Gaiagan J., Gardyna S.,
RA      Grant G., Hagos B., Heaford A., Horton L., Howland J.C., Johnson R.,
RA      Jones C., Kann L., Karatas A., Klein J., Lehoczy J., Liu C., Locke K.,
RA      MacDonald P., Marguis N., McEwan P., McGurk A., McKernan K., McLaughlin J.,
RA      Melidim J., Morrow J., Naylor J., Norman C.H., O'Connor T., O'Donnell P.,
RA      Peterson K., Pollara V., Riley R., Roy A., Santos R., Severy P.,
RA      Strange-Thomann N., Stojanovic N., Subramanian A., Talamas J., Testaye S.,
RA      Tirrell A., Vassiliev H., Vo A., Wheeler J., Wu X., Wymann D., Ye W.J.,
RA      Zimmer A., Zody M.;
RT
RL      Submitted (16-OCT-1999) to the EMBL/GenBank/DBJ databases.
RL      Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL      Cambridge, MA 02141, USA
XX
On Feb 28, 2000 this sequence version replaced gl:6479009.
CC      All repeats were identified using RepeatMasker:
CC      Smit, A.F.A. & Green, P. (1996-1997)
CC      http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC      ----- Genome Center
CC      Center: Whitehead Institute/ MIT Center for Genome Research
CC      Center code: WtBR
CC      Web site: http://www-seq.wi.mit.edu
CC      Contact: sequence.submissions@genome.wi.mit.edu
CC      ----- Project Information
CC      Center project name: U3520
CC      Center clone name: 16_D_24
CC      ----- Summary Statistics
CC      Sequencing vector: M13; M77815; 100% of reads
CC      Chemistry: Dye-terminator Big Dye; 100% of reads
CC      Assembly program: Phrap; version 0.960731
CC      Consensus quality: 133371 bases at least 040
CC      Consensus quality: 152498 bases at least 030
CC      Consensus quality: 159565 bases at least 020
CC      Insert size: 170000; agarose-fp
CC

```

```
CC Insert size: 163889; sum-of-ctrls
CC Quality coverage: 3.9 in Q20 bases; agarose-fp
CC Quality coverage: 3.9 in Q20 bases; sum-of-ctrls
CC -----
CC * NOTE: This is a 'working draft' sequence. It currently
CC * consists of 19 contigs. The true order of the pieces
CC * is not known and their order in this sequence record is
CC * arbitrary. Gaps between the contigs are represented as
CC * runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence
CC * as soon as it is available and the accession number will
CC * be preserved.
CC 1
CC * 1358: contlg of 1358 bp in length
CC * 1359 1458: gap of 100 bp
CC * 1459 3377: contlg of 1919 bp in length
CC * 3378 3477: gap of 100 bp
CC * 3478 6249: contlg of 2772 bp in length
CC * 6250 6349: gap of 100 bp
CC * 6350 9888: contlg of 3539 bp in length
CC * 9889 9988: gap of 100 bp
CC * 9989 12785: contlg of 2797 bp in length
CC * 12786 12885: gap of 100 bp
CC * 12886 16977: contlg of 4092 bp in length
CC * 16978 17077: gap of 100 bp
CC * 17078 22033: contlg of 4956 bp in length
CC * 22034 22133: gap of 100 bp
CC * 22134 26597: contlg of 4464 bp in length
CC * 26598 26697: gap of 100 bp
CC * 26698 31119: contlg of 4422 bp in length
CC * 31120 31219: gap of 100 bp
CC * 31220 35831: contlg of 4612 bp in length
CC * 35832 35931: gap of 100 bp
CC * 35932 40025: contlg of 4094 bp in length
CC * 40026 40125: gap of 100 bp
CC * 40126 48025: contlg of 7900 bp in length
CC * 48026 48125: gap of 100 bp
CC * 48126 57672: contlg of 9547 bp in length
CC * 57673 57772: gap of 100 bp
CC * 57773 68763: contlg of 10991 bp in length
CC * 68764 68863: gap of 100 bp
CC * 68864 82748: contlg of 13885 bp in length
CC * 82749 82848: gap of 100 bp
CC * 82849 96756: contlg of 13908 bp in length
CC * 96757 96856: gap of 100 bp
CC * 96857 118840: contlg of 21984 bp in length
CC * 118841 118940: gap of 100 bp
CC * 118941 142301: contlg of 23361 bp in length
CC * 142302 142401: gap of 100 bp
CC * 142402 165689: contlg of 23288 bp in length.
XX
FH Key Location/Qualifiers
FH 1.165689
FH source
FH /db_xref:"taxon:9606"
FH /organism:"Homo sapiens"
FH /clone:"RP11-16D24"
FH /clone.lib:"RPC1-11 Human Male BAC"
FT misc_feature
FT 1.1358
FT /note:"assembly-fragment"
FT misc_feature
FT 1459.3377
FT /note:"assembly-fragment"
FT misc_feature
FT 3478.6249
FT /note:"assembly-fragment"
FT misc_feature
FT 6350.9888
FT /note:"assembly-fragment"
FT misc_feature
FT 9989.12785
FT /note:"assembly-fragment"
FT misc_feature
FT 12886.16977
FT /note:"assembly-fragment"
FT misc_feature
FT 17078.22033
FT /note:"assembly-fragment"
FT misc_feature
FT 22134.26597
FT /note:"assembly-fragment-clone_end:17-vector_slide:right"
```

```
FT misc_feature
FT 26698.31119
FT /note:"assembly-fragment-clone_end:sp6-vector_slide:right"
FT misc_feature
FT 31220.35831
FT /note:"assembly-fragment"
FT misc_feature
FT 35932.40025
FT /note:"assembly-fragment"
FT misc_feature
FT 40126.48025
FT /note:"assembly-fragment"
FT misc_feature
FT 48126.57672
FT /note:"assembly-fragment"
FT misc_feature
FT 57773.68763
FT /note:"assembly-fragment"
FT misc_feature
FT 68864.82748
FT /note:"assembly-fragment"
FT misc_feature
FT 82849.96756
FT /note:"assembly-fragment"
FT misc_feature
FT 96857.118840
FT /note:"assembly-fragment"
FT misc_feature
FT 118941.142301
FT /note:"assembly-fragment"
FT misc_feature
FT 142402.165689
FT /note:"assembly-fragment"
XX
SQ Sequence 165689 BP; 48440 A; 34083 C; 32848 G; 48513 T; 1805 other;
+
alignment_scores:
Quality: 53.00 Length: 14
Ratio: 4.818 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 64.286
alignment_block:
US-09-471-276-831_COPY_1_16 x AC011936/rev ..
Align seg 1/1 to reverse of: AC011936 from: 1 to: 165689
3 MetLeuValAlaPheLeuLeuTrpGlyValThrTrpGly 16
:::||||| ||||| ||||| |||||
146811 GTGCTTTCGCAATTCTTCACTTTCGCTGCTTTCGCGG 146770
seq_name: gb_hug:AC013707
seq_documentation_block:
LOCUS AC013707 166349 bp DNA HTG 09-MAR-2000
DEFINITION Homo sapiens clone RP11-22L6, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
ACCESSION AC013707.2 GI:7212036
VERSION AC013707.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 166349)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-22L6
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 166349)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckertly,R., Boguslavsky,L., Boukhvalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Deatellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gargyala,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karalas,A., Klein,J.,
Lehoczky,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McQuirk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tittel,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
```

JOURNAL Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 9, 2000 this sequence version replaced g1:12583848.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Genome Center

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L4143

Center clone name: 22.L.6

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 163375 bases at least Q40

Consensus quality: 164944 bases at least Q30

Consensus quality: 165450 bases at least Q20

Insert size: 164000; agarose-fp

Insert size: 165749; sum-of-contrigs

Quality coverage: 9.2 in Q20 bases; agarose-fp

Quality coverage: 9.1 in Q20 bases; sum-of-contrigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contrigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contrigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1114: contrig of 1114 bp in length
 * 1115 1214: gap of 100 bp
 * 1215 2270: contrig of 1056 bp in length
 * 2271 2370: gap of 100 bp
 * 2371 3468: contrig of 1098 bp in length
 * 3469 3568: gap of 100 bp
 * 3569 4585: contrig of 1017 bp in length
 * 4586 4685: gap of 100 bp
 * 4686 28060: contrig of 23375 bp in length
 * 28061 28160: gap of 100 bp
 * 28161 79056: contrig of 50896 bp in length
 * 79057 79156: gap of 100 bp
 * 79157 166349: contrig of 87193 bp in length.
 Location/Qualifiers

FEATURES

source

1. 166349

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="RP11-22L6"

/clone_lib="RP11-11 Human Male BAC"

1. 1114

/note="assembly-fragment"

1215..2270

/note="assembly-fragment"

2371..3468

/note="assembly-fragment"

3569..4585

/note="assembly-fragment"

4686..28060

/note="assembly-fragment"

28161..79056

/note="assembly-fragment"

79157..166349

/note="assembly-fragment"

vector_side:left"

misc_feature

51366 a 31182 c 30388 g 52813 t 600 others

BASE COUNT

ORIGIN

alignment_scores:

Quality: 53.00 Length: 10
 Ratio: 5.300 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 70.000

alignment_block:

US-09-471-276-831_COPY_1_16 x AC013707 ..

Align seg 1/1 to: AC013707 from: 1 to: 166349

6 ValpheleuLeuTrpGlyValThrTrp 15

146806 CTTTTCCTAAATTCGTGGGGAATTACATGC 146835

seq_name: gb_hlg:AC024359

seq_documentation_block:

LOCUS AC024359 168208 bp DNA

DEFINITION Homo sapiens chromosome 8 clone RP11-272P15 map 8, WORKING DRAFT

SEQUENCE, 5 unordered pieces.

ACCESSION AC024359

VERSION AC024359.3 GI:14336589

KEYWORDS HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 168208)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

JOURNAL Homo sapiens chromosome 8, clone RP11-272P15

REFERENCE 2 (bases 1 to 168208)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,

Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,

Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,

Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,

Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hages,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,U.J., Landers,T., Laroque,K., Lehotzky,J., Levine,R.,

Lieu,C., Liu,G., Locke,K., Macdonald,P., Margus,N., McCarthy,M.,

McEwan,J., McGov,A., McKernan,K., McPheters,R., Melgrim,J.,

Menes,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J., Naylor,D.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Ollivar,T.M.,

Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C.,

Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,

Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tittel,A.,

Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,

Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and

Zody,M.

Direct Submission

Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 11, 2001 this sequence version replaced g1:12583848.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Genome Center

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L5830

Center clone name: 272.P.15

Summary Statistics

Sequencing vector: M13; M77815; 5% of reads

Sequencing vector: Plasmid; n/a; 95% of reads

```

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165277 bases at least Q40
Consensus quality: 166822 bases at least Q30
Consensus quality: 167378 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 167808; sum-of-ctrls
Quality coverage: 5.7 in Q20 bases; agarose-fp
Quality coverage: 5.7 in Q20 ba.
NOTE: This is a 'working draft' sequence. It currently
consists of 5 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
*
* 1 85799: contig of 85799 bp in length
* 85800 85899: gap of 100 bp
* 85900 96614: contig of 10715 bp in length
* 96615 96714: gap of 100 bp
* 96715 116316: contig of 19602 bp in length
* 116317 116416: gap of 100 bp
* 116417 141745: contig of 25329 bp in length
* 141746 141845: gap of 100 bp
* 141846 168208: contig of 26363 bp in length.
Location/Qualifiers
FEATURES
source
1.168208
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/chromosome:"8"
/map:"8"
/clone:"RP11-272P15"
/clone.lib:"RP11 Human Male BAC"
1.85799
/note:"assembly-fragment"
clone_end:SP6
vector_side:left"
misc_feature
85900..96614
/note:"assembly-fragment"
96715..116316
/note:"assembly-fragment"
116417..141745
/note:"assembly-fragment"
141846..168208
/note:"assembly-fragment"
clone_end:T7
vector_side:right"
BASE COUNT 52087 a 31238 c 30497 g 53985 t 401 others
ORIGIN
alignment_scores:
Quality: 53.00 Length: 10
Ratio: 5.300 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 70.000
alignment_block:
US-09-471-276-831_COPY_1_16 x AC024359 ..
Align seq 1/1 to: AC024359 from: 1 to: 168208
6 ValPhcLeuLeuLeuTrrpGlyValThrTrp 15
:::|||||:|||||:|||||:|||||:|||||:
142177 CTTTTCATAATCTGTCGGCAATACATGCG 142206
seq_name: gb_hvg:AC023569
seq_documentation_block:
LOCUS AC023569 169607 bp DNA HTG 06-APR-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-513M1 map 18, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
ACCESSION AC023569

```

```

VERSION
AC023569.2 GI:7466913
KEYWORDS
HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 169607)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-513M1
Unpublished
2 (bases 1 to 169607)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Bozislavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campiloro,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Darellano,K., Dekar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karalas,A.,
Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marguis,N., McCarthy,M.,
McEwan,P., McQuirk,A., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mlyova,T., Miranda,C., Mlenya,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triggillo,J., Vassiliev,H., Vtel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zahoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 6, 2000 this sequence version replaced gi:5978265.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/MW/repeatmasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 513_M1
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158843 bases at least Q40
Consensus quality: 163699 bases at least Q30
Consensus quality: 166127 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 168007; sum-of-ctrls
Quality coverage: 4.4 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-ctrls
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 17 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
*
* 1 1030: contig of 1030 bp in length
* 1031 1130: gap of 100 bp
* 1131 2617: contig of 1487 bp in length
* 2618 2717: gap of 100 bp
* 2718 5109: contig of 2392 bp in length
* 5110 5209: gap of 100 bp

```

```

* 5210 8324: contig of 3115 bp in length
* 8325 8424: gap of 100 bp
* 8425 11014: contig of 2590 bp in length
* 11015 11114: gap of 100 bp
* 11115 13394: contig of 2280 bp in length
* 13395 13494: gap of 100 bp
* 13495 17701: contig of 4207 bp in length
* 17702 17801: gap of 100 bp
* 17802 22017: contig of 4216 bp in length
* 22018 22117: gap of 100 bp
* 22118 25803: contig of 3686 bp in length
* 25804 25903: gap of 100 bp
* 25904 34269: contig of 8366 bp in length
* 34270 34369: gap of 100 bp
* 34370 44782: contig of 10413 bp in length
* 44783 44882: gap of 100 bp
* 44883 58944: contig of 14062 bp in length
* 58945 59044: gap of 100 bp
* 59045 73159: contig of 14115 bp in length
* 73160 73259: gap of 100 bp
* 73260 88014: contig of 14755 bp in length
* 88015 88114: gap of 100 bp
* 88115 103661: contig of 15547 bp in length
* 103662 103761: gap of 100 bp
* 103762 134075: contig of 30314 bp in length
* 134076 134175: gap of 100 bp
* 134176 169607: contig of 35432 bp in length.
Location/Qualifiers
1. 169607
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-513M1"
/clone_id="RPC1-11 Human Male BAC"
1. 1030
/note="assembly_fragment"
1131. 2617
/note="assembly_fragment"
/note="assembly_fragment"
2718. 5109
/note="assembly_fragment"
5210. 8324
/note="assembly_fragment"
8425. 11014
/note="assembly_fragment"
11115. 13394
/note="assembly_fragment"
13495. 17701
/note="assembly_fragment"
17802. 22017
/note="assembly_fragment"
22118. 25803
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
25904. 34269
/note="assembly_fragment"
34370. 44782
/note="assembly_fragment"
44883. 58944
/note="assembly_fragment"
59045. 73159
/note="assembly_fragment"
73260. 88014
/note="assembly_fragment"
88115. 103661
/note="assembly_fragment"
clone_end:r7
vector_side:right"
103762. 134075
/note="assembly_fragment"
134176. 169607
/note="assembly_fragment"

```

```

BASE COUNT 50874 a 32780 c 33552 g 50796 t 1605 others
ORIGIN
alignment_scores:
    Quality: 53.00      Length: 15
    Ratio: 4.818        Gaps: 0
    Percent Similarity: 73.333    Percent Identity: 60.000
alignment_block:
US-09-471-276-831_COPY_1_16 x AC023569
Align seg 1/1 to: AC023569 from: 1 to: 169607
2 SetMetLeuValAlpHeLeuLeuEutRpglyValThTrpgly 16
166598 TCATACCTGTATATCTTCCTACGTGGAGAACGACGAGTGGGCT 166642
seq_name: gb_hlg:AC092580
seq_documentation_block:
LOCUS AC092580 172005 bp DNA HTG 19-JUL-2001
DEFINITION Homo sapiens chromosome 2 clone RP11-16D24, WORKING DRAFT SEQUENCE,
5 unordered pieces.
ACCESSION AC092580 AC011936
VERSION AC092580.1 GI:14916165
KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Jul 19, 2001 this sequence version replaced gi:7107944.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0016D24
Drafting center: WIBR
----- Summary Statistics -----
Sequencing vector: M13: 37%
Sequencing vector: Plasmid: 63%
Chemistry: Dye-primer ET; 2% of reads
Chemistry: Dye-terminator Big Dye; 98% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171007 bases at least Q40
Consensus quality: 171655 bases at least Q20
Consensus quality: 171787 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 173588; sum-of-contigs
Quality coverage: 9.32 in Q20 bases; sum-of-contigs
Quality coverage: 9.86 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

*      1      1453: contig of 1453 bp in length
*      1553: gap of unknown length
*      1554: 2880: contig of 1327 bp in length
*      2881: 2980: gap of unknown length
*      2981: 7450: contig of 4470 bp in length
*      7451: 7550: gap of unknown length
*      7551: 67265: contig of 59715 bp in length
*      67266: 67365: gap of unknown length
*      67366: 172005: contig of 104640 bp in length.

FEATURES
    source
        1..172005
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="2"
            /clone="RP11-16D24"

    misc_feature
        1..1453
            /note="assembly_name:Contig17"

    misc_feature
        1554..2880
            /note="assembly_name:Contig21"

    misc_feature
        2981..7450
            /note="assembly_name:Contig22"

    clone_end::T7
    vector_side::right"

    misc_feature
        7551..67265
            /note="assembly_name:Contig23"

    misc_feature
        67366..172005
            /note="assembly_name:Contig24"

BASE COUNT      51161 a 34632 c 35587 g 50224 t      401 others
ORIGIN

alignment_scores:
    Quality: 53.00      Length: 14
    Ratio: 4.818      Caps: 0
    Percent Similarity: 78.571      Percent Identity: 64.266

alignment_block:
US-09-471-276-831_COPY_1_16 x AC092580 ..

Align seg 1/1 to: AC092580 from: 1 to: 172005

3 MetLeuValValPheLeuLeuTgPgiValThTgPgly 16
GGCTCTTCAATTCTTAAGCTTGGCTGTTCTTGGGCGG 39859

seq_name: gb_pr:AC016080

seq_documentation_block:
LOCUS      AC016080      176921 bp      DNA      PRI      30-MAY-2001
DEFINITION Homo sapiens, clone RP11-23N14, complete sequence.
ACCESSION      AC016080
VERSION
AC016080.5      GI:14017515
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 176921)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
Homo sapiens, clone RP11-23N14
Unpublished
2 (bases 1 to 176921)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,D., Barna,N., Beckerly,R., Boguslavsky,L., Bouknighter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Croke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J.C., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Liu,C., Locke,K., Macdonald,P., Marquis,A., Klein,J.,
McEwan,P., McGurk,A., McKernan,K., McDonald,J., McLaughlin,J., Meisid,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,

```

```

TITLE
JOURNAL
COMMENT
REFERENCE
AUTHORS

Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tjirell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 176921)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Bouknighter,B., Brown,A.,
Camarata,J., Campobiano,A., Chang,J., Chongel,Y., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., McPheters,R., Melgrim,J., Menais,L.,
Milnova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,K., Seaman,S.,
Severy,P., Sougnez,C., Spencer,B., Strange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Travers,M., Travis,N., Trillio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zemek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-MAY-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 10, 2001 this sequence version replaced gi:13357421.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Center code: MIBR
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4401
Center clone name: 23_N_14
-----
FEATURES
    source
        1..176921
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="RP11-23N14"
            /clone_lib="RPCI-11 Human Male BAC"

    repeat_region
        68..156
            /rpt_family="MIR"

    repeat_region
        658..679
            /rpt_family="AT-rich"

    repeat_region
        complement(703..911)
            /rpt_family="MIR"

    repeat_region
        complement(1027..1416)
            /rpt_family="MLT1A2"

    repeat_region
        1495..1549
            /rpt_family="MLT-1nt"

    repeat_region
        1551..1591
            /rpt_family="CA)n"

    repeat_region
        complement(1594..2024)
            /rpt_family="MLT2B1"

    repeat_region
        complement(2440..2751)
            /rpt_family="A1ub"

    repeat_region
        complement(2756..3641)
            /rpt_family="LMC1"

    repeat_region
        3742..3821
            /rpt_family="TA)n"

    repeat_region
        4084..4214
            /rpt_family="L2"

    repeat_region
        complement(4505..4712)

```



```

repeat_region /rpt_family="HERVL" complement(4774..5479)
repeat_region /rpt_family="PTR5" complement(5488..5766)
repeat_region /rpt_family="HERVL" complement(6391..7000)
repeat_region /rpt_family="HERVL" 7360..8042
repeat_region /rpt_family="L2" complement(8077..8416)
repeat_region /rpt_family="THEIA" 8558..8976
repeat_region /rpt_family="MER4D" complement(8977..9268)
repeat_region /rpt_family="Alusg" 9269..9724
repeat_region /rpt_family="MER4D" 9725..9988
repeat_region /rpt_family="L2" 10200..10245
repeat_region /rpt_family="MLT2B1" 10246..11011
repeat_region /rpt_family="MER4E" 11012..11265
repeat_region /rpt_family="MLT2B1" 11086..11181
unsure /note="<30 qual SNGL region" 11126..11163
unsure /note="single clone coverage" 11222..11345
unsure /note="single clone coverage" 11262..11296
repeat_region /rpt_family="MLT2B1" 11762..11862
repeat_region /rpt_family="MLT2B2" complement(12099..12225)
repeat_region /rpt_family="Charlieta" complement(12226..12515)
repeat_region /rpt_family="AluY" complement(12516..13835)
repeat_region /rpt_family="Charlieta" 15559..15694
repeat_region /rpt_family="CT-rich" complement(15716..15771)
repeat_region /rpt_family="L1PB1" 15772..16017
repeat_region /rpt_family="L1PB1" 16018..16038
repeat_region /rpt_family="AT-rich" complement(16258..16848)
repeat_region /rpt_family="L2" 16941..16964
repeat_region /rpt_family="(CAAA)n" 17119..17148
repeat_region /rpt_family="AT-rich" complement(17644..18010)
repeat_region /rpt_family="L1M4" 18487)
repeat_region /rpt_family="L1MDa" 18589..18615
repeat_region /rpt_family="AT-rich" 18938..19306
repeat_region /rpt_family="MLT1K" 19362..19676
repeat_region /rpt_family="Alusx" 20057..20081
repeat_region /rpt_family="AT-rich" complement(20153..20224)
repeat_region /rpt_family="L1ME3" complement(20394..20635)
repeat_region /rpt_family="L1M4" complement(20718..20874)
repeat_region /rpt_family="L1M4"

```

```

repeat_region complement(20927..21384)
repeat_region /rpt_family="L1MEc" 22737..22790
repeat_region /rpt_family="AT-rich" 22875..22905
repeat_region /rpt_family="AT-rich" 23424..23501
repeat_region /rpt_family="MIR" 24358..24385
repeat_region /rpt_family="AT-rich" complement(24715..24880)
repeat_region /rpt_family="MIR" complement(25053..25141)
repeat_region /rpt_family="MIR" complement(25150..25331)
repeat_region /rpt_family="L1PA13" complement(25883..25978)
repeat_region /rpt_family="MIR" complement(26056..27011)
repeat_region /rpt_family="L1M4" complement(28909..28989)
repeat_region /rpt_family="L2" complement(30023..30235)
repeat_region /rpt_family="MIR" 30337..30860
repeat_region /rpt_family="trigger7" 30867..30954
repeat_region /rpt_family="trigger7" 31873..31896

alignment_scores:
    quality: 53.00    length: 10
    ratio: 5.300      gaps: 0
    percent similarity: 100.000    percent identity: 70.000

alignment_block:
US-09-471-276-831_COPY_1_16 x AC016080 ..
Align seg 1/1 to: AC016080 from: 1 to: 176921
6 ValPheLeuLeuLeuTrpGlyValThrTrp 15
:::|||||:|||||:|||||:|||||
111951 CTTTTCATATCTGTGGCAATTACATCG 111980

seq_name: gb_pl:ATCHRIV40
seq_documentation_block:
LOCUS ATCHRIV40 197405 bp DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 40.
ACCESSION AL161540
VERSION AL161540.2 GI:7268224
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
leemke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de,Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK.
E-mail: michael.bevan@bsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/chal/
this fragment has an overlap with ATCHRIV39 at the 5' end and an
overlap with ATCHRIV41 at the 3' end.
Location/Qualifiers
1..197405
FEATURES
SOURCE

```

```

/organism."Arabidopsis thaliana"
/variety."Columbia"
/db_xref."taxon:3702"
/chromosome."4"
complement(8150..8280)
/gene."AT4g14790"
/number.1
8150..10349
/gene."AT4g14790"
complement(join(8150..8280,8363..8531,8621..8698,
8884..9006,9126..9246,9299..9539,9618..9739,9883..10162,
10286..10349))
/gene."AT4g14790"
complement(join(8150..8280,8363..8531,8621..8698,
8884..9006,9126..9246,9299..9539,9618..9739,9883..10162,
10286..10349))
/gene."AT4g14790"
/note."similarity to probable RNA helicase SUV3,
ATP-dependent, yeast (Saccharomyces cerevisiae)
contains ATP/GTP-binding site motif A (P-loop)
AA136-143:EF-hand calcium-binding domain AA383-395"
/codon_start.1
/product."RNA helicase like protein"
/protein_id."CAB78521.1"
/db_xref."GI:7268225"
/translation."MAYSVVRLRKVSALGISRVLDSELNLYELCSFYICQVSRRL
DMFVVSVDGEFLNVSADGSLWRPFEEPEFGDLLRGVLTNRNKSQSPKDFPG
TGTSKDFPDLTGPHWYPLARKKKRVLLHVGPTNSGKYSALKHLEOSSGCVG
PLRLAMEVAKRLNKANVPCDLITGQERDLVEGTHKAVYEMADVTSYDCAITDEL
QASLARLMLKSTRFCLGFATRALGIADEHLICGDPVAVPLVEDLLKVTDDVE
VHYTERLSPVPLKVPSSVSIKTGDLVTFESKRDYAYAKKTERAGKHLCSVYGS
LPETRTAQATRFNDETFDVLVSDAIGMGLNLISRIIFSTLQKYDGSFTDLTV
SEIKQIAGRAGRFQSKFPGCEVTCIHKEDPLHLHSLSKSPSPILEANLMH"
complement(8281..8362)
/number.1
complement(8363..8531)
/gene."AT4g14790"
/number.2
complement(8532..8620)
/number.2
complement(8621..8698)
/gene."AT4g14790"
/number.3
complement(8699..8883)
/number.3
complement(8884..9006)
/gene."AT4g14790"
/number.4
complement(9007..9125)
/number.4
complement(9126..9246)
/gene."AT4g14790"
/number.5
complement(9247..9298)
/number.5
complement(9299..9539)
/gene."AT4g14790"
/number.6
complement(9540..9617)
/number.6
complement(9618..9739)
/gene."AT4g14790"
/number.7
complement(9740..9882)
/number.7
complement(9883..10162)
/gene."AT4g14790"
/number.8
complement(10163..10285)
/number.8
complement(10286..10349)
/gene."AT4g14790"
/number.9

```

```

gene
10918..12578
/gene."AT4g14800"
10918..11075
/gene."AT4g14800"
/number.1
join(10918..11075,11167..11290,12261..12578)
/gene."AT4g14800"
/note."strong similarity to proteasome chain Hsc7-1
protein, human
contains EST gb:A1993573.1, Y09359, R65485, T22127,
F13874"
/codon_start.1
/product."proteasome chain protein"
/protein_id."CAB78522.1"
/db_xref."GI:7268226"
/translation."MECVFGLVGNMFAIVADTSAVHSILHKRKEDKIMTLDSHKLY
AASGPDROVFTFYQKNVSLYQFRNGIPLSTAAANFTGELATALARKRPYSNLT
MAGYDKEGASLYITDYIATLHKYDKGAFGYFSLSITMDRHYRSDMSVEALELVD
KCLIEIRSLVIAAPNPIKIVDKGAREYCWIRISTADA"
11076..11166
/gene."AT4g14800"
/number.1
11167..11290
/gene."AT4g14800"
/number.2
11291..12260
/gene."AT4g14800"
/number.2
12261..12578
/gene."AT4g14800"
/number.3
12579..13042
/gene."AT4g14810"
/number.1
12928..15983
/gene."AT4g14810"
join(12928..13042,13394..13501,13559..13702,15165..15253,
15348..15365,15846..15983)
/gene."AT4g14810"
/note."similarity to Fun12p, Saccharomyces cerevisiae
contains EST gb:AA042489, A1998638.1, T88290"
/codon_start.1
/product."hypothetical protein"
/protein_id."CAB78523.1"
/db_xref."GI:7268227"
/translation."MMIVDGNOKFLIRSEFYKCRFGSGGGGAGVAVSDLRKN
QSRSIQRESKDRLPEKITEVSGARNGGSDVKKARREHDELRTSONRRRCCLL
RGGYGGGGEGEDREYEDENAVSYLRTPPKTPYQIKPINSNNH1KMNSQDSNPNVYL
DGGSGQLGINVNEQALALPKACHVETPPASRCHSKINSLSMH"
13001..13024
13043..13393
/gene."AT4g14810"
/number.1
13394..13501
/gene."AT4g14810"
/number.2
13502..13558
/gene."AT4g14810"
/number.2
13559..13702
/gene."AT4g14810"
/number.3
13703..15164
/gene."AT4g14810"
/number.3
15165..15253
/gene."AT4g14810"
/number.4
15254..15347
/gene."AT4g14810"
/number.4
15348..15365
/gene."AT4g14810"

```

```
intron      /number=5
            /15366..15845
            /gene="ATg14810"
exon        /number=5
            /15846..15983
            /gene="ATg14810"
            /number=6
            /18314..20558
            /gene="ATg14820"
            /complement(18314..19925)
            /gene="ATg14820"
            /number=1
            /complement(join(18314..19925,20002..20558))
            /gene="ATg14820"
            /complement(join(18314..19925,20002..20558))
            /gene="ATg14820"
            /note="similarity to 130 KD LEUCINE-RICH PROTEIN (LRP 130)
            (GP130) L130_HUMAN
            Contains Cytochrome c family heme-binding site signature
            Aa682-687:Prokaryotic membrane lipoprotein lipid
            attachment site Aa252-262:Prokaryotic membrane lipoprotein
            lipid attachment site Aa344-354:Prokaryotic membrane
            lipoprotein lipid attachment site Aa513-523
            contains EST gb:AA585978, F15442, F15441"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="CA878524.1"
            /db_xref="GI:7268228"

alignment_scores:
    Quality: 53.00      Length: 11
    Ratio: 4.818      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 63.636

alignment_block:
US-09-471-276-831_COPY_1_16 x ATCHRIV40 ..
Align seg 1/1 to: ATCHRIV40 from: 1 to: 197405

5 ValValPheLeuLeuLeuTrpGlyValThrTrp 15
:::|||||:::|||||:::|||||
135990 ATCTTATTTGTTACTTGGGCGATGACATCG 136022

seq_name: gb_pl:ATFCA2

seq_documentation_block:
LOCUS   ATFCA2      202860 bp      DNA      PLN      29-JUN-1999
DEFINITION   Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
            No. 2.
ACCESSION   297337
VERSION     297337
KEYWORDS    GI:5302774
SOURCE      '
ORGANISM    '
            thale cress.
            Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
            Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
            1 (bases 1 to 202860)
            Bevan, M., Stiekema, W., Murphy, G., Wambutt, R., Pohl, T., Terry, N.,
            Kreis, M., Kavanagh, T., Entian, K. D., Rieger, M., James, R.,
            Pujdomenech, P., Hatzopoulos, P., Obermaier, B., Duesterhoft, A.,
            Jones, J., Palme, K., Ansoorge, W., Delseny, M., Bancroft, I.,
            Meves, H. W., Schellier, C. and Chalhatzis, N.
            Unpublished
            2 (bases 1 to 202860)
            EU Arabidopsis sequencing project.
            Direct Submission
            Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schellier@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7JY Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
```

```
COMMENT
On Jun 30, 1999 this sequence version replaced gi:2244829.
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATFCA1 at the 5' end and an
overlap with ATFCA3 at the 3' end.

FEATURES
            source
            location/Qualifiers
            1..202860
            /organism="Arabidopsis thaliana"
            /variety="Columbia"
            /db_xref="taxon:3702"
            /chromosome="4"
            1..70917
            /organism="Arabidopsis thaliana"
            /db_xref="taxon:3702"
            /clone="BAC TAMU2D4"
            13673..13975
            /gene="d13405w"
            13673..13975
            /gene="d13405w"
            /number=1
            13673..13975
            /gene="d13405w"
            /note="weak similarity to N-methyl-D-aspartate receptor
            glutamate-binding chain, rat"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="CAB10252.1"
            /db_xref="GI:2244830"
            /db_xref="SPTREMBL:O23329"
            /translation="MAKSDIEFGGNELYPCMKSSSELRMAPIRKIYSLISLQILYV
            GVSAYVVRPIPERPIETHRKGLAVFVILLPLRIYSEFILLFIFHFLCSTIYA"
            /complement(join(15633..15841,16271..16328,16682..16731,
            16815..17057,17098..17157,17295..17739,18223..18333,
            18701..19012))
            /gene="d13410c"
            /complement(15633..15841)
            /gene="d13410c"
            /number=1
            15633..19012
            /gene="d13410c"
            /complement(join(15633..15841,16271..16328,16682..16731,
            16815..17057,17098..17157,17295..17739,18223..18333,
            18701..19012))
            /gene="d13410c"
            /note="weak similarity to DNA-binding protein"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="CAB10253.1"
            /db_xref="GI:2244831"
            /db_xref="SPTREMBL:O23330"
            /translation="WEKLMVPTWRPDVRYRPPEPLPEMEFLARSMVSALAEVSKALT
            PPNOIILSKTEEEDEEEDPISVVDGDDTDTGLVNGNPFSAFSETSQVMDRIIS
            HSOEVSPTSGRLSHSGPLNGSLSDSPVSPESDDIKQCFRANKNSLVNSOFRS
            TAAVPGPITATATOSKTYRWMLKDKREKKKEETRAHMQIIAASVAGVAAVAIAA
            ATAASSCGKDEQAKTMDAASAATTIAACQVEAAEYMGAEKREYLAVSAAVNRVS
            AGDITMTLAGAATIDDSIDLGISKLELEISIRLGRVOTLAKARAKVWNIYASVTPM
            DKGLSTGSSNNVNGSSSSSSSHSGLVQDENLGICSRSEMLARGGELLRKRGD
            LHWKIVSYIKMKNOVMKMSRHVGGFYTKKKNIYDVIVIKNPAMPNGRHLLLEGDD
            LRYGLKIVMGDVEFEVKSOREYEMWTQGVSRLLVLAAEKRFK"
            /complement(15842..16270)
            /number=1
            /complement(16271..16328)
            /gene="d13410c"
            /number=2
            /complement(16329..16681)
            /number=2
            /complement(16682..16731)
            /gene="d13410c"
            /number=3
            /complement(16732..16814)
            /number=3
            /complement(16815..17057)
```



```
repeat_region complement(34085..34439)
repeat_region /rpt_family="THEIB"
repeat_region 34573..34845
repeat_region /rpt_family="THEIB"
repeat_region 34846..35412
repeat_region /rpt_family="THEIB-int"
repeat_region 36413..36768

alignment_scores:
  Quality: 53.00 Length: 10
  Ratio: 5.300 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 70.000

alignment_block:
US-09-471-276-831_COPY_1_16 x AC012410/rev ..
Align seg 1/1 to reverse of: AC012410 from: 1 to: 220715

6 ValPheLeuLeuTrpGlyValThrTrp 15
:::|||||:::|||||:::|||||
87840 CTTTCTNATCTCTGCGGATTCATG 87811

seq_name: gb_in:CET27E9

seq_documentation_block:
LOCUS CET27E9 22245 bp DNA INV 20-JUN-2001
DEFINITION Caenorhabditis elegans cosmid T27E9, complete sequence.
ACCESSION Z82059
VERSION Z82059.1 GI:1914550
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 2245);
none.
Genome sequence of the nematode C. elegans: a platform for
lifestigling biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
JOURNAL MEDLINE The C.elegans Sequencing Consortium.
99069613
REMARK 2 (bases 1 to 2245)
REFERENCE LLOYD,C.R.
AUTHORS Direct Submission
JOURNAL Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1HQ, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
On Mar 30, 1997 this sequence version replaced gi:1666020.
Coding sequences below are predicted from computer analysis, using
predictions from GeneFINDER (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye-terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone T27E9.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone T27E9 is at 19189 in
sequence AL032665.
The true right end of clone T27E9 is at 16930 in
sequence Z92813.
The start of this sequence (1..104) overlaps with the end of
sequence AL032665.
The end of this sequence (22142..22245) overlaps with the start of
sequence Z92813.
For a graphical representation of this sequence and its analysis
see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
name=T27E9.
Location/Qualifiers
```

```
source
1..22245
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="T27E9"
complement(424..651)
/gene="T27E9.2"
complement(424..651)
/gene="T27E9.2"
/note="predicted using GeneFINDER
Similarity to bovine ubiquinol-cytochrome c reductase
complex 11 KD protein (SW:P00126)
CDNA EST YK268b2.5 comes from this gene"
/codon_start=1
/protein_id="CAB04873.1"
/db_xref="GI:3880275"
/db_xref="SPTREMBL:O45864"
/translation="MSSHKEDPLADVDQLNQYREKADHYTERKSLIDECNDHYNSR
SNTETCTCHQEMADYVHHLDHCAMPKAFSLK"
join(1298..1355,1464..1528,1574..1794,1848..2406)
/gene="T27E9.1"
join(1298..1355,1464..1528,1574..1794,1848..2406)
/gene="T27E9.1"
/note="predicted using GeneFINDER
Similarity to bovine ADP/ATP carrier protein (SW:P32007),
contains similarity to Pfam domain: PF00153 (Mitochondrial
carrier proteins). Score=382.1, E-value=1.8e-111, N=1
CDNA EST EMBL:Z14700 comes from this gene: CDNA EST
EMBL:M8914 comes from this gene: CDNA EST EMBL:Z14645
comes from this gene
CDNA EST YK9b5.5 comes from this gene: CDNA EST
EMBL:T00728 comes from this gene: CDNA EST CEMSE14F comes
from this gene
CDNA EST CEMSE77F comes from this gene: CDNA EST YK73c5.5
comes from this gene: CDNA EST YK78d5.3 comes from this
gene
CDNA EST YK119c4.3 comes from this gene: CDNA EST YK60b7.3
comes from this gene: CDNA EST EMBL:Z14655 comes from this
gene
CDNA EST EMBL:Z14688 comes from this gene: CDNA EST
EMBL:Z14689 comes from this gene: CDNA EST YK9b5.3 comes
from this gene
CDNA EST CEMSA46R comes from this gene: CDNA EST CEMSE77B
comes from this gene: CDNA EST CEMSE14R comes from this
gene
CDNA EST EMBL:T00927 comes from this gene: CDNA EST
CEMSB32R comes from this gene: CDNA EST EMBL:T01509 comes
from this gene
CDNA EST CEMSA46F comes from this gene: CDNA EST CEMSB32FB
comes from this gene: CDNA EST CEMSB55F comes from this
gene
CDNA EST CEMSG56F comes from this gene: CDNA EST YK62a3.3
comes from this gene: CDNA EST YK67e1.7 comes from this
gene
CDNA EST YK67f9.3 comes from this gene: CDNA EST YK75e5.3
comes from this gene: CDNA EST YK75d7.5 comes from this
gene
CDNA EST YK96e10.3 comes from this gene: CDNA EST
YK104f5.3 comes from this gene: CDNA EST YK60b7.5 comes
from this gene
CDNA EST YK62e3.5 comes from this gene: CDNA EST YK62b6.5
comes from this gene: CDNA EST YK64g12.5 comes from this
gene
CDNA EST YK67e1.5 comes from this gene: CDNA EST YK67f9.5
comes from this gene: CDNA EST YK75d7.5 comes from this
gene
CDNA EST YK75e5.5 comes from this gene: CDNA EST YK75e7.5
comes from this gene: CDNA EST YK82f9.5 comes from this
gene
CDNA EST YK84d7.5 comes from this gene: CDNA EST YK89b8.5
comes from this gene: CDNA EST YK90c1.5 comes from this
gene
CDNA EST YK96e10.5 comes from this gene: CDNA EST
```

```

yk104f5.5 comes from this gene; cDNA EST yk107a1.5 comes
from this gene
cDNA EST yk111f3.5 comes from this gene; cDNA EST
yk115g5.5 comes from this gene; cDNA EST yk56f6.3 comes
from this gene
cDNA EST yk94b5.3 comes from this gene; cDNA EST
yk100h10.3 comes from this gene; cDNA EST yk55d8.5 comes
from this gene
cDNA EST yk56b1.5 comes from this gene; cDNA EST yk56f6.5
comes from this gene; cDNA EST yk58a5.5 comes from this
gene
cDNA EST yk70c2.5 comes from this gene; cDNA EST yk94b5.5
comes from this gene; cDNA EST yk99h4.5 comes from this
gene
cDNA EST yk100h10.5 comes from this gene; cDNA EST
yk119c4.5 comes from this gene; cDNA EST yk172c2.3 comes
from this gene
cDNA EST yk172g6.3 comes from this gene; cDNA EST yk77b3.5
comes from this gene; cDNA EST yk77g11.5 comes from this
gene
cDNA EST yk108f10.5 comes from this gene; cDNA EST
yk166e11.5 comes from this gene; cDNA EST yk166c7.5 comes
from this gene
cDNA EST yk172c2.5 comes from this gene; cDNA EST
yk172g6.5 comes from this gene; cDNA EST yk175a4.5 comes
from this gene
cDNA EST yk150b7.5 comes from this gene; cDNA EST
yk150b7.3 comes from this gene; cDNA EST yk453h1.3 comes
from this gene
cDNA EST yk453h1.5 comes from this gene; cDNA EST
yk293d11.3 comes from this gene; cDNA EST yk293d11.5 comes
from this gene
cDNA EST yk292d10.3 comes from this gene; cDNA EST
yk292d10.5 comes from this gene; cDNA EST yk291c7.3 comes
from this gene
cDNA EST yk291c7.5 comes from this gene; cDNA EST
yk231e9.3 comes from this gene; cDNA EST yk231e9.5 comes
from this gene
cDNA EST yk201f4.3 comes from this gene; cDNA EST
yk201f4.5 comes from this gene; cDNA EST yk291b10.5 comes
from this gene
cDNA EST yk563b5.3 comes from this gene; cDNA EST
yk671b2.3 comes from this gene; cDNA EST yk563b5.5 comes
from this gene
cDNA EST yk648h4.5 comes from this gene; cDNA EST
yk66f2.3 comes from this gene; cDNA EST yk66f2.5 comes
from this gene
cDNA EST yk671b2.5 comes from this gene"
/codon_start-1
/protein_id:"CAB04874.1"
/db_xref:"GI:3880276"
/db_xref:"SPTREMBL:O45865"
/translation:"MSKESRDYKKFLIDLASGTAANVSTNAPIERVKLLQYVD
ASKAIADVDRKICIMDVLIIRPKDQGAALMRGLAVIRFPQANFARQDYKAI
FLGLDKRDKDFWKFAGNLASGAAGATSLCFVPLPFAIRLADIGKADREFKGL
ADCLIKIKSDPGIGLGRGPFVSVOGIIIRAAAFGMDPRKAVFASDQKLNFFAM
GIAOVTVGSGILSYPMDFVRRMMQSGRKDILYKNTLPDCAKRIQNEGSAEFKA
LSNFRGTGGLAIATYDEIOKFL"
complement(join(2885..2971,3024..3104,3505..3635,
3691..4144,4190..4278,4326..4362))
/gene:"T27E9.3"
complement(join(2885..2971,3024..3104,3505..3635,
3691..4144,4190..4278,4326..4362))
/gene:"T27E9.3"
complement(join(2885..2971,3024..3104,3505..3635,
3691..4144,4190..4278,4326..4362))
/gene:"T27E9.3"
/notice:"Similarity to Xenopus cell division protein kinase
5 (SW:PS1166), contains similarity to Pfam domain: PF00069
(Eukaryotic protein kinase domain), Score:279.8,
E-value:1.1e-80, N=1
cDNA EST yk101f9.3 comes from this gene
cDNA EST yk101f9.5 comes from this gene
cDNA EST yk171d8.5 comes from this gene
cDNA EST yk171d8.3 comes from this gene
cDNA EST yk367b1.3 comes from this gene

```

```

cDNA EST yk367b1.5 comes from this gene
cDNA EST yk388b11.3 comes from this gene
cDNA EST yk549a4.3 comes from this gene
cDNA EST yk379e6.5 comes from this gene
cDNA EST yk388b11.5 comes from this gene
cDNA EST yk549a4.5 comes from this gene"

alignment_scores:
  Quality: 52.00      Length: 16
  Ratio: 4.000       Gaps: 0
  Percent Similarity: 81.250   Percent Identity: 43.750

alignment_block:
US-09-471-276-831_COPY_1_16 x CET27E9
Align seg 1/1 to: CET27E9 from: 1 to: 22245
1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
::: ::: ||||| ::::::::::::::::::::|||
13617 CTTCTGCTGCACAGTCTCGGAATCGTGTGGGCGATCATCGTGGCGC 13664

seq_name: gb_in:AC084471
seq_documentation_block:
LOCUS AC084471 95539 bp DNA INV 04-NOV-2000
DEFINITION Caenorhabditis briggsae cosmid CB048320, complete sequence.
ACCESSION AC084471
VERSION AC084471.1 GI:11094921
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis briggsae.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 95539)
REFERENCE
AUTHORS Washington University Genome Sequencing Center.
TITLE The C. briggsae Genome Sequencing Project
UNPUBLISHED
REFERENCE
AUTHORS 2 (bases 1 to 95539)
TITLE Waterston, R.
JOURNAL Direct Submission
Submitted (04-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jspliel@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.
Location/Qualifiers
source
1..95539
/organism:"Caenorhabditis briggsae"
/strain:"GujaRat G16"
/db_xref:"taxon:6238"
/clone:"CB048320"
BASE COUNT 29558 a 17793 c 18212 g 29976 t
ORIGIN

alignment_scores:
  Quality: 52.00      Length: 12
  Ratio: 4.333       Gaps: 0
  Percent Similarity: 100.000   Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x AC084471
Align seg 1/1 to: AC084471 from: 1 to: 95539

```


101072 CTTTATCTGTTCTTACTGTGAGAGTTCATGG 101110

seq_name: gb_htg:AC068151

seq_documentation_block:

LOCUS AC068151 187399 bp DNA HTG 04-JUN-2000

DEFINITION Homo sapiens chromosome 2 clone RP11-310D24 map 2, WORKING DRAFT

SEQUENCE, 24 unordered pieces.

ACCESSION AC068151.2 GI:8247821

VERSION AC068151.2 HTG: HTGS-PHASE1; HTGS-DRAFT.

KEYWORDS

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 187399)

Biren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 2, clone RP11-310D24

Unpublished

2 (bases 1 to 187399)

Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,

Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,

Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S.,

Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,

Gallagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,

Klein, J., Lafocque, K., Lamazares, R., Landers, T., Lehoczy, J.,

Leavitt, R., Liu, G., Locke, K., Macdonald, P., Margulis, N.,

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,

Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J.,

Murphy, T., Naylor, T., Norman, C.H., O'Connor, T., O'Donnell, P.,

O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Plante, N.,

Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,

Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Testa, S., Theodore, J., Tirrell, A., Travers, M., Triggiani, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,

Young, G., Zalnoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (29-APR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 4, 2000 this sequence version replaced gi:7670145.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L10091

Center clone name: 310_D24

Summary Statistics

Sequencing vector: M13; W7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 177725 bases at least Q40

Consensus quality: 181810 bases at least Q30

Consensus quality: 183611 bases at least Q20

Insert size: 187000; agarose-gel

Insert size: 185099; sum-of-coverage

Quality coverage: 4.9 in Q20 bases; sum-of-coverage

Quality coverage: 4.9 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently

consists of 24 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      1109: contig of 1109 bp in length
*      1110 1209: gap of 100 bp
*      1210 2333: contig of 1124 bp in length
*      2334 2433: gap of 100 bp
*      2434 5417: contig of 2984 bp in length
*      5418 5517: gap of 100 bp
*      5518 6201: contig of 684 bp in length
*      6202 6301: gap of 100 bp
*      6302 9249: contig of 2948 bp in length
*      9250 9349: gap of 100 bp
*      9350 12316: contig of 2967 bp in length
*      12317 12416: gap of 100 bp
*      12417 14957: contig of 2541 bp in length
*      14958 15057: gap of 100 bp
*      15058 19294: contig of 4237 bp in length
*      19295 19394: gap of 100 bp
*      19395 21704: contig of 2310 bp in length
*      21705 21804: gap of 100 bp
*      21805 25038: contig of 3234 bp in length
*      25039 25138: gap of 100 bp
*      25139 29374: contig of 4236 bp in length
*      29375 29474: gap of 100 bp
*      29475 34094: contig of 4620 bp in length
*      34095 34194: gap of 100 bp
*      34195 40194: contig of 6000 bp in length
*      40195 40294: gap of 100 bp
*      40295 45768: contig of 5474 bp in length
*      45769 45868: gap of 100 bp
*      45869 50773: contig of 4905 bp in length
*      50774 50873: gap of 100 bp
*      50874 55763: contig of 4890 bp in length
*      55764 55863: gap of 100 bp
*      55864 61980: contig of 6117 bp in length
*      61981 62080: gap of 100 bp
*      62081 66132: contig of 4052 bp in length
*      66133 66232: gap of 100 bp
*      66233 72920: contig of 6688 bp in length
*      72921 73020: gap of 100 bp
*      73021 81047: contig of 8027 bp in length
*      81048 81147: gap of 100 bp
*      81148 88415: contig of 7268 bp in length
*      88416 88515: gap of 100 bp
*      88516 99641: contig of 11126 bp in length
*      99642 99741: gap of 100 bp
*      99742 135048: contig of 35307 bp in length
*      135049 135148: gap of 100 bp
*      135149 187399: contig of 52251 bp in length.

```

FEATURES

source

```

1..187399 Homo sapiens
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone_lib="RP11-310D24"
/clone_11b="RP11-11 Human Male BAC"
1..1109
/clone="assembly_fragment"
/clone_11b="assembly_fragment"
1210..2333
/clone="assembly_fragment"
2434..5417
/clone="assembly_fragment"
5518..6201
/clone="assembly_fragment"
6302..9249
/clone="assembly_fragment"
9350..12316
/clone="assembly_fragment"
12417..14957
/clone="assembly_fragment"

```

```
misc_feature /note:"assembly-fragment"
15058..19294
/note:"assembly-fragment"
misc_feature 19395..21704
/note:"assembly-fragment"
clone_end:r7
vector_side:right"
misc_feature 21805..25038
/note:"assembly-fragment"
25139..29374
/note:"assembly-fragment"
misc_feature 29475..34094
/note:"assembly-fragment"
misc_feature 34195..40194
/note:"assembly-fragment"
misc_feature 40295..45768
/note:"assembly-fragment"
misc_feature 45869..50773
/note:"assembly-fragment"
misc_feature 50874..55763
/note:"assembly-fragment"
misc_feature 55864..61980
/note:"assembly-fragment"
misc_feature 62081..66132
/note:"assembly-fragment"
misc_feature 66233..72920
/note:"assembly-fragment"
misc_feature 73021..81047
/note:"assembly-fragment"
misc_feature 81148..88415
/note:"assembly-fragment"
misc_feature 88516..99641
/note:"assembly-fragment"
misc_feature 99742..135048
/note:"assembly-fragment"
misc_feature 135149..187399
/note:"assembly-fragment"
BASE COUNT 52163 a 41609 c 40858 g 50466 t 2303 others
ORIGIN
Alignment-scores:
Quality: 52.00 Length: 14
Ratio: 4.727 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 57.143
alignment_block:
US-09-471-276-831_COPY_1_16 x AC068151/rev ..
Align seg 1/1 to reverse of: AC068151 from: 1 to: 187399
3 MotLeuValValPheLeuLeuTrrpGlyValThrTrpGly 16
||||| ||| :::::::::::::::::::: |||||
100370 ATGCTTGAGGCTCTGATGATCGTCGCGACTGCACTGGGGA 100329
seq_name: gb_hcg:AC009146
seq_documentation_block:
LOCUS AC009146 188249 bp DNA HTG 26-JAN-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-557E6, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC009146
VERSION AC009146.2 GI:6758890
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 188249)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 188249)
```

```
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jan 26, 2000 this sequence version replaced gi:5685929.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
*****
* NOTE: This record contains 105 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 753 1044: contig of 292 bp in length
* gap of unknown length
* 1045 1553: contig of 509 bp in length
* gap of unknown length
* 1554 2201: contig of 648 bp in length
* gap of unknown length
* 2202 2459: contig of 258 bp in length
* gap of unknown length
* 2460 3128: contig of 669 bp in length
* gap of unknown length
* 3129 3735: contig of 607 bp in length
* gap of unknown length
* 3736 4365: contig of 630 bp in length
* gap of unknown length
* 4366 5162: contig of 797 bp in length
* gap of unknown length
* 5163 5767: contig of 605 bp in length
* gap of unknown length
* 5768 6395: contig of 628 bp in length
* gap of unknown length
* 6396 7140: contig of 745 bp in length
* gap of unknown length
* 7141 7673: contig of 533 bp in length
* gap of unknown length
* 7674 8427: contig of 754 bp in length
* gap of unknown length
* 8428 8535: contig of 108 bp in length
* gap of unknown length
* 8536 9297: contig of 762 bp in length
* gap of unknown length
* 9298 9783: contig of 486 bp in length
* gap of unknown length
* 9784 10486: contig of 703 bp in length
* gap of unknown length
* 10487 11259: contig of 773 bp in length
* gap of unknown length
* 11260 11441: contig of 182 bp in length
* gap of unknown length
* 11442 11658: contig of 217 bp in length
* gap of unknown length
* 11659 12804: contig of 1146 bp in length
* gap of unknown length
* 12805 13783: contig of 979 bp in length
* gap of unknown length
* 13784 14983: contig of 1200 bp in length
* gap of unknown length
* 14984 15042: contig of 59 bp in length
* gap of unknown length
* 15043 15968: contig of 926 bp in length
* gap of unknown length
```

```

* 15969 16759: contig of 791 bp in length
* 16760 16909: contig of 150 bp in length
* 16910 17075: contig of 166 bp in length
* 17076 17651: contig of 576 bp in length
* 17652 18442: contig of 791 bp in length
* 18443 19359: contig of 917 bp in length
* 19360 19637: contig of 278 bp in length
* 19638 19845: contig of 208 bp in length
* 19846 20057: contig of 212 bp in length
* 20058 20794: contig of 737 bp in length
* 20795 21514: contig of 720 bp in length
* 21515 22109: contig of 595 bp in length
* 22110 22821: contig of 712 bp in length
* 22822 23715: contig of 894 bp in length
* 23716 24809: contig of 1094 bp in length
* 24810 25572: contig of 763 bp in length
* 25573 25751: contig of 179 bp in length
* 25752 25802: contig of 51 bp in length
* 25803 26709: contig of 907 bp in length
* 26710 27649: contig of 940 bp in length
* 27650 28561: contig of 912 bp in length
* 28562 30107: contig of 1546 bp in length
* 30108 31238: contig of 1131 bp in length
* 31239 32178: contig of 940 bp in length
* 32179 32246: contig of 68 bp in length
* 32247 33821: contig of 1575 bp in length
* 33822 34752: contig of 931 bp in length
* 34753 35666: contig of 914 bp in length
* 35667 36541: contig of 875 bp in length
* 36542 37859: contig of 1318 bp in length
* 37860 38724: contig of 865 bp in length
* 38725 40403: contig of 1679 bp in length
* 40404 41689: contig of 1286 bp in length
* 41690 42617: contig of 928 bp in length
* 42618 43987: contig of 1370 bp in length
* 43988 45759: contig of 1772 bp in length
* 45760 47312: contig of 1553 bp in length

```

```

* 47313 48547: contig of 1235 bp in length
* 48548 50263: contig of 1716 bp in length
* 50264 51618: contig of 1355 bp in length
* 51619 53515: contig of 1897 bp in length
* 53516 55107: contig of 1592 bp in length
* 55108 56444: contig of 1337 bp in length
* 56445 58112: contig of 1668 bp in length
* 58113 59998: contig of 1886 bp in length
* 59999 61454: contig of 1456 bp in length
* 61455 62356: contig of 902 bp in length
* 62357 64123: contig of 1767 bp in length
* 64124 65418: contig of 1295 bp in length
* 65419 67216: contig of 1798 bp in length
* 67217 69587: contig of 2371 bp in length
* 69588 71418: contig of 1831 bp in length
* 71419 74182: contig of 2764 bp in length
* 74183 76376: contig of 2194 bp in length
* 76377 78385: contig of 2009 bp in length
* 78386 81113: contig of 2728 bp in length
*
alignment_scores:
    Quality: 52.00      Length: 14
    Ratio: 4.727       Gaps: 0
    Percent Similarity: 78.571    Percent Identity: 57.143

alignment_block:
US-09-471-276-831_COPY_1_16 x AC009146/rev ..
Align seg 1/1 to reverse of: AC009146 from: 1 to: 188249

      3 MetLeuValIvalPheLeuLeuLeuTrpGlyValThrTrpGly 16
      169024 ATGCTTGAGTCGTCGATGATCTCTGGGACTGACGACGGGGA 168983

seq_name: gb_hhg:AC023766

seq_documentation_block:
LOCUS AC023766 227611 bp DNA HTG 10-MAY-2000
DEFINITION Homo sapiens chromosome 5 clone RP11-459A22 map 5, WORKING DRAFT
SEQUENCE, 62 unordered pieces.
ACCESSION AC023766
VERSION AC023766.2 GI:7767814
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 227611)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 5, clone RP11-459A22
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 227611)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

```

Anderson, S., Baldwin, J., Barne, N., Beda, F., Boguslavsky, L.,
Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeRellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,
Fenster, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J., C. Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Landers, T., Larcocque, K., Lehotsky, J., Levine, R.,
Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M.,
McKwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J.,
Mensu, L., Milnova, T., Miranda, C., Mienga, V., Morrow, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M.,
Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A.,
Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilsson, B.,
Wu, X., Wyman, D., Ye, W. J., Young, G., Zalnoun, J., Zimmer, A. and
Zody, M.

TITLE
JOURNAL

COMMENT

Submitted (17-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 10, 2000 this sequence version replaced g1:6984473.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L6351

Center clone name: 459_A.22

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 161281 bases at least Q40

Consensus quality: 183874 bases at least Q30

Consensus quality: 202702 bases at least Q20

Insert size: 162000; agarose-fp

Insert size: 221511; sum-of-contigs

Quality coverage: 4.9 in Q20 bases; agarose-fp

Quality coverage: 3.6 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1178: contig of 1178 bp in length
* 1179 1278: gap of 100 bp
* 1279 2325: contig of 1047 bp in length
* 2326 2425: gap of 100 bp
* 2426 3504: contig of 1079 bp in length
* 3505 3604: gap of 100 bp
* 3605 4812: contig of 1208 bp in length
* 4813 4912: gap of 100 bp
* 4913 6336: contig of 1424 bp in length
* 6337 6436: gap of 100 bp
* 6437 7606: contig of 1170 bp in length
* 7607 7706: gap of 100 bp
* 7707 9220: contig of 1514 bp in length
* 9221 9320: gap of 100 bp
* 9321 10404: contig of 1084 bp in length
* 10405 10504: gap of 100 bp
* 10505 11787: contig of 1283 bp in length
* 11788 11887: gap of 100 bp

11888 13396: contig of 1509 bp in length
* 13397 13496: gap of 100 bp
* 13497 14524: contig of 1028 bp in length
* 14525 14624: gap of 100 bp
* 14625 15925: contig of 1301 bp in length
* 15926 16025: gap of 100 bp
* 16026 16427: contig of 402 bp in length
* 16428 16527: gap of 100 bp
* 16528 17984: contig of 1457 bp in length
* 17985 18084: gap of 100 bp
* 18085 19657: contig of 1773 bp in length
* 19658 19957: gap of 100 bp
* 19958 21461: contig of 1504 bp in length
* 21462 21561: gap of 100 bp
* 21562 23157: contig of 1596 bp in length
* 23158 23257: gap of 100 bp
* 23258 24526: contig of 1269 bp in length
* 24527 24626: gap of 100 bp
* 24627 25955: contig of 1329 bp in length
* 25956 26055: gap of 100 bp
* 26056 27976: contig of 1921 bp in length
* 27977 28076: gap of 100 bp
* 28077 29674: contig of 1798 bp in length
* 29675 29974: gap of 100 bp
* 29975 31504: contig of 1530 bp in length
* 31505 31604: gap of 100 bp
* 31605 32976: contig of 1372 bp in length
* 32977 33076: gap of 100 bp
* 33077 34237: contig of 1161 bp in length
* 34238 34337: gap of 100 bp
* 34338 35654: contig of 1317 bp in length
* 35655 35754: gap of 100 bp
* 35755 37631: contig of 1877 bp in length
* 37632 37731: gap of 100 bp
* 37732 39294: contig of 1563 bp in length
* 39295 39394: gap of 100 bp
* 39395 41364: contig of 1970 bp in length
* 41365 41464: gap of 100 bp
* 41465 43336: contig of 1872 bp in length
* 43337 43436: gap of 100 bp
* 43437 45488: contig of 2052 bp in length
* 45489 45588: gap of 100 bp
* 45589 47902: contig of 2314 bp in length
* 47903 48002: gap of 100 bp
* 48003 50923: contig of 2921 bp in length
* 50924 51023: gap of 100 bp
* 51024 53227: contig of 2204 bp in length
* 53228 53327: gap of 100 bp
* 53328 56721: contig of 3394 bp in length
* 56722 56821: gap of 100 bp
* 56822 58718: contig of 1897 bp in length
* 58719 58818: gap of 100 bp
* 58819 61916: contig of 3098 bp in length
* 61917 62016: gap of 100 bp
* 62017 64899: contig of 2883 bp in length
* 64900 64999: gap of 100 bp
* 65000 69476: contig of 4477 bp in length
* 69477 69576: gap of 100 bp
* 69577 74433: contig of 4857 bp in length
* 74434 74533: gap of 100 bp
* 74534 78765: contig of 4232 bp in length
* 78766 78865: gap of 100 bp
* 78866 83783: contig of 4918 bp in length
* 83784 83883: gap of 100 bp
* 83884 88554: contig of 4671 bp in length
* 88555 88654: gap of 100 bp
* 88655 92887: contig of 4233 bp in length
* 92888 92987: gap of 100 bp
* 92988 97174: contig of 4187 bp in length
* 97175 97274: gap of 100 bp
* 97275 101602: contig of 4328 bp in length
* 101603 101702: gap of 100 bp
* 101703 107534: contig of 5832 bp in length

```

* 107535 107634: gap of 100 bp
* 107635 112107: contig of 4473 bp in length
* 112108 112207: gap of 100 bp
* 112208 117172: contig of 4965 bp in length
* 117173 117272: gap of 100 bp
* 117273 121060: contig of 3788 bp in length
* 121061 121160: gap of 100 bp
* 121161 124846: contig of 3686 bp in length
* 124847 124946: gap of 100 bp
* 124947 129670: contig of 4724 bp in length
* 129671 129770: gap of 100 bp
* 129771 134148: contig of 4378 bp in length
* 134149 134248: gap of 100 bp
* 134249 141225: contig of 6977 bp in length
* 141226 141325: gap of 100 bp
* 141326 148226: contig of 6901 bp in length
* 148227 148326: gap of 100 bp
* 148327 154101: contig of 5775 bp in length
* 154102 154201: gap of 100 bp
* 154202 160592: contig of 6391 bp in length
* 160593 160692: gap of 100 bp
* 160693 170151: contig of 9459 bp in length
* 170152 170251: gap of 100 bp
* 170252 179670: contig of 9419 bp in length
* 179671 179770: gap of 100 bp
* 179771 189052: contig of 9282 bp in length
* 189053 189152: gap of 100 bp
* 189153 200238: contig of 11086 bp in length
* 200239 200338: gap of 100 bp
* 200339 213317: contig of 12879 bp in length
* 213318 213317: gap of 100 bp
* 213318 227611: contig of 14294 bp in length.

```

FEATURES

```

Location/Qualifiers
1..227611
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5"

```

```

alignment_scores:
  Quality: 52.00      Length: 9
  Ratio: 5.778        Gaps: 0
Percent Similarity: 100.000    Percent Identity: 88.889

```

alignment_block:

US-09-471-276-831_COPY_1_16 x AC023766 ..

Align seg 1/1 to: AC023766 from: 1 to: 227611

```

8 LeuLeuLeuTrpGlyValThrTrpGly 16
IIIIIIIIIIIIIIIIIIIIIIIIIIIIII
83676 CTCTCTCTGTGGGCGCCACATGCGGC 83702

```

seq_name: gb_hhg:AC023509

seq_documentation_block:

```

LOCUS      AC023509 173188 bp      DNA      HTG      19-JUL-2001
DEFINITION Homo sapiens chromosome 12 clone RP11-793H13, WORKING DRAFT
SEQUENCE   5 unordered pieces.
ACCESSION  AC023509
VERSION    AC023509.31 GI:14861644
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     human.
ORGANISM   Homo sapiens
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 173188)
Muzny,D.M., Adams,C., Adio-Odola,B., Altschman,F.R., Allen,C.,
Alstorkors,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blumh,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Blevins,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flang,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,B., Hernandez,O., Hodgson,A., Hughes,M., Hollway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Joliver,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Krtovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokemwo,S., Oguh,M., Okunodu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,C., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Woodson,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Glibos,R.

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project information
Center project name: HNGA
Center clone name: RP11-793H13
Sequencing vector: Plasmid; M77789
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 4% of reads
Chemistry: Dye-terminator Big Dye: 96% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 170543 bases at least Q40
Consensus quality: 171067 bases at least Q30
Consensus quality: 171311 bases at least Q20
Estimated insert size: 172224; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 10.4x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 5 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

```

* 1 57577: contig of 57577 bp in length
* 57578 57677: gap of unknown length
* 57678 125390: contig of 67713 bp in length
* 125391 125490: gap of unknown length
* 125491 145187: contig of 19697 bp in length
* 145188 145287: gap of unknown length
* 145288 166917: contig of 21630 bp in length
* 166918 167017: gap of unknown length
* 167018 173188: contig of 6171 bp in length.
FEATURES
  source
    1..173188
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="12"
    /clone="RP11-793H13"
BASE COUNT 47157 a 38496 c 38905 g 48222 t 408 others
ORIGIN
alignment_scores:
  Quality: 51.50 Length: 13
  Ratio: 4.682 Gaps: 1
  Percent Similarity: 84.615 Percent Identity: 76.923
alignment_block:
US-09-471-276-831_COPY_1_16 x AC023509/rev ..
Align seg 1/1 to reverse of: AC023509 from: 1 to: 173188
5 ValValpheteuleuleutpicyValthr...Ttgcly 16
|||||
51803 GTTTCCTTTTACTTGGGATACACGCGTGGCGG 51765
seq_name: gb_hcg:AC068889
seq_documentation_block:
LOCUS AC068889 182930 bp DNA HTG 26-JUN-2001
DEFINITION Homo sapiens chromosome 12 clone RP11-774122, WORKING DRAFT
SEQUENCE 12 unordered pieces.
AC068889
AC068889 27 GT:14547453
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.
SOURCE
  human.
ORGANISM
  Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 182930)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alldredge,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bovle,S., Briteva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burich,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,T., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Doultwaite,K.J., Draper,H.,
Duran-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Franze,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Gatz,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Hawlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homel,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivel,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kralovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsgaard,H., Lozardo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,B., Massey,E., Mashiney,E., McLeod,M.P., Meador,M.,
McGill,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,R.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S.,
Oguth,M., Okwunonu,G., Oragunye,N., Oviado,R., Pace,A., Payton,B.,
Peery,J., Perenz,L., Peters,L., Pickens,R., Prins,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojiboken,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtat,N.,
Slison,I., Sodergren,E., Sonalka,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Swalek,A., Tabor,P., Tamejisa,A., Tamejisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 182930)
Worley,K.C.
Direct Submission
Submitted (11-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 25, 2001 this sequence version replaced gi:13605936.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBDZ
Center clone name: RP11-774122
----- Summary Statistics
Sequencing vector: Plasmid; M7789
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye 3 of reads
Chemistry: Dye-terminator Big Dye 3 of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 180190 bases at least Q40
Consensus quality: 185089 bases at least Q30
Consensus quality: 187580 bases at least Q20
Estimated insert size: 181412; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 9.7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
37419: contig of 37419 bp in length
37420 37519: gap of unknown length
37520 59469: contig of 21950 bp in length
59470 59569: gap of unknown length
59570 82549: gap of 22980 bp in length
82550 82649: gap of unknown length
82650 105815: contig of 23166 bp in length
105816 105915: gap of unknown length
105916 123596: contig of 17681 bp in length
123597 123696: gap of unknown length
123697 142064: contig of 18368 bp in length
142065 142164: gap of unknown length
142165 152808: contig of 10644 bp in length
152809 152908: gap of unknown length
152909 155523: contig of 12615 bp in length
155524 155623: gap of unknown length
155624 172843: contig of 7222 bp in length
172845 172945: gap of unknown length
172946 177159: contig of 4214 bp in length
177160 177259: gap of unknown length

```


* 177260 180098: contig of 2839 bp in length
* 180099 180198: gap of unknown length
* 180199 182930: contig of 2732 bp in length.
Location/Qualifiers
1. 182930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-774I22"
BASE COUNT 48339 a 43236 c 42404 g 47843 t 1108 others
ORIGIN

alignment_scores:
Quality: 51.50 Length: 13
Ratio: 4.682 Gaps: 1
Percent Similarity: 84.615 Percent Identity: 76.923

alignment_block:
US-09-471-276-831_COPY_1_16 x AC068889 ..
Align seg 1/1 to: AC068889 from: 1 to: 182930

5 ValValPheLeuLeuLeuTrrpGlyValThr...TrrpGly 16
|||||
43155 GTTTCCTTTTACTTGTGGGGATACCTGGGTGGGG 43193

seq_name: gb_ro:RNO302031

seq_documentation_block:
LOCUS RNO302031 1741 bp mRNA ROD 03-JUL-2001
DEFINITION Rattus norvegicus mRNA for putative alpha 1B-glycoprotein (ORF1).
ACCESSION AJ302031
VERSION AJ302031.1 GI:11877347
KEYWORDS alpha 1B-glycoprotein.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1741)
Gardmo, C., Persson, B. and Mode, A.
Cloning of a novel growth hormone-regulated rat complementary
deoxyribonucleic acid with homology to the human
alpha1B-glycoprotein, characterizing a new protein family
Endocrinology 142 (6), 2695-2701 (2001)
21255654
2 (bases 1 to 1741)
Gardmo, C.
Direct Submission
Submitted (12-DEC-2000) Gardmo C., Department of Medical Nutrition,
Karolinska Institutet, Novum F60, Huddinge, S-14186, SWEDEN
Location/Qualifiers
1. 1741
/organism="Rattus norvegicus"
/strain="Sprague Dawley"
/db_xref="taxon:10116"
/sex="female"
/tissue_type="liver"
/dev_stage="adult"
30. 1571
/note="ORF1"
/codon_start=1
/product="putative alpha 1B-glycoprotein"
/protein_id="CAC19029.1"
/db_xref="GI:11877348"
/db_xref="SPTREMBL:Q9EPH1"
/translation="MSLTITVILMGFTLPGNGLNLNSGSPPELRAPOSILPEKAN
LTLVCAVDLPTVFELINGMFLISOVRLTEPVLSRSLGATTSNNGYRCRCYEP
PDIOLPALSKMTLSNALEVGKEPLPPSAHADVSWITPGCLPVTMCVRAMGV
TYLRLKEGVDTOKPDVOKHGTAGELIYRPGYSCSYLTHAGSESEPSAIYTIKMSA
TOLPSLCLMGSYLTIYPOKTHETTLACKAPRNAAEFOLRQGERVLNIQGFSPTRDTI
YYVNLKELDNOSPFCRYRMHYMHVWSEDSKPVLELMSDEKLPAVLVLAEPSSNMLE

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
CDS

PGSTVQLCTAHKAGLRFGLQROGKPDLVVOMLNSSGTEAVEFLHNIISTIDSGNYS
IYMEQADPFSSASSSEPLELRINGPAPRPLEALMKGVPLGHEAIFQCHGVPRVSM
ELVREGEKTPPMWASTSTSFNLKLSVCGQHTONYSCTRTALSPTFESGIDPVEV
VVEGS"
99
/note="translation change from Asn to Asp"
/replace="g"
1692
BASE COUNT 450 a 458 c 430 g 403 t
ORIGIN

variation
polyA-site
BASE COUNT 450 a 458 c 430 g 403 t
ORIGIN

alignment_scores:
Quality: 51.00 Length: 16
Ratio: 3.923 Gaps: 0
Percent Similarity: 81.250 Percent Identity: 62.500

alignment_block:
US-09-471-276-831_COPY_1_16 x RNO302031 ..
Align seg 1/1 to: RNO302031 from: 1 to: 1741

1 MetSerMetLeuValValPheLeuLeuLeuTrrpGlyValThrTrrpGly 16
|||||
30 ATGCTCTGTTGACTGACTGACTGCTCTGCGGTTTCACCTCGGGC 77

seq_name: gb_pat:AX045627

seq_documentation_block:
LOCUS AX045627 1908 bp DNA PAT 24-NOV-2000
DEFINITION Sequence 7 from Patent WO0066787.
ACCESSION AX045627
VERSION AX045627.1 GI:11344029
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1908)
Kopchick, J. and Tjong, J.
Growth hormone-regulatable liver genes and proteins, and uses
thereof
Patent: WO 0066787-A 7 09-NOV-2000;
Ohio University (US)
Location/Qualifiers
1. 1908
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 525 a 475 c 477 g 431 t
ORIGIN

alignment_scores:
Quality: 51.00 Length: 16
Ratio: 3.923 Gaps: 0
Percent Similarity: 81.250 Percent Identity: 62.500

alignment_block:
US-09-471-276-831_COPY_1_16 x AX045627 ..
Align seg 1/1 to: AX045627 from: 1 to: 1908

1 MetSerMetLeuValValPheLeuLeuLeuTrrpGlyValThrTrrpGly 16
|||||
11 ATGCTCTGCTGCGCTGACTGCTGCTGCGGTTTCACCTCGGGC 58

seq_name: gb_ro:AF269193

seq_documentation_block:
LOCUS AF269193 2933 bp mRNA ROD 21-JUN-2000
DEFINITION Mus musculus bromodomain-containing FSH-like protein FSRG2 (FSRG2)
mRNA, complete cds.
ACCESSION AF269193

[illegible]

```

REFERENCE          Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS            Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
TITLE              Direct Submission
JOURNAL            Nigro, V.
REFERENCE          Submitted (12-JUL-1997) Nigro V., Istituto di Patologia Generale e
AUTHORS            Oncologia, Seconda Università degli Studi di Napoli, Larghetto
TITLE              S. Antonio a Caponapoli, 2, NAPOLI, 80138, ITALY
JOURNAL            2 (bases 1 to 3628)
FEATURES           Belisto A., Piluso, G. and Nigro, V.
SOURCE             A novel X-linked calpain-like protease
                  Unpublished
                  location/Qualifiers
                    1..3628
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="X"
                      /map "q2"
                      183..2108
                      /codon_start=1
                      /product="calpain-like protease, CANPX"
                      /protein_id="CAA04051.1"
                      /db_xref="GI:2274962"
                      /translation="MGPKLFRKNOKYOELKQECIKRSLFCPTPLPNDLSLFRTL
LPCKVVKRRPODTCDDPHLVGNTSNHQLTGRGLKIPMYSASFCLAVOSHTKTIP
NMHEQWDPQTEREYAGIFPRFMHPMEETEVYLDLLPTINDLVFSTSMNEFWNN
ALKEAYAKLGLCGEALDGALTIDIIVDFGTLAETVDOMKRGYTELVEERYKLFGEL
KYRTTKGCLICCSIESPSOEEOEVEETDMGLKGTYMTDIRIRLGERLVEVSARK
VVYWRRLNPLGRBMSGPMSEIISEWOULTASDRKNLGLYMSDGEFMSLEDFCRNF
HKLNCRNVNVPITFGRELESYSVGCTPVDDPLMBNSGCCYNADPTLQROYLTYP
EDGKRATMSIQDKDLRTYRRNRGRDNITIGFLFKVYMKRFRIHLHYTOERGSTYIP
IDRTKVALSKYLRKKGNVLVPTMFQHONGETSEFLRIFSEVPVOLRELTLDMPMSCWN
LARCPYRVTVQITVHSADELKRYANETVNPYLVIKCGKEVSPVQKNVVAIFDRO
AIFRRTTDIPILIVGVNNSRKFCDOPLGGVTLDADPSDCDLKSLYLRRKGGPYAKVK
QGHSFVVISDDTEL"
BASE COUNT         1004 a      926 c      809 g      889 t
ORIGIN
alignment_scores:
    Quality:       51.00      Length:      8
    Ratio:         6.375     Gaps:        0
Percent Similarity: 100.000   Percent Identity: 100.000
Alignment_block:
US-09-471-276-831_COPY_1_16 x HSCANPX/rev ..
Align seg 1/1 to reverse of: HSCANPX from: 1 to: 3628
          9 LeuLeuTrpGlyValThrTrpGly 16
          |||
          2726 TTACTGTGGGGGTGACCTGGGGT 2703
seq_name: gb_htg:AC019542
seq_documentation_block:
LOCUS               AC019542      45605 bp      DNA      HTG      03-JAN-2000
DEFINITION          Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered
                     pieces.
ACCESSION            AC019542
VERSION              AC019542.1  GI:665355
KEYWORDS             HTG; HTGS; PHASE2.
SOURCE               fruit fly.
ORGANISM             Drosophila melanogaster
                     Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                     Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                     Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE            Adams, M. and Venter, J.C.
AUTHORS              Direct Submission
TITLE                Submitted (30-DEC-1999) Celera Genomics, 45 West Guile Drive,
JOURNAL              Rockville, MD, USA
COMMENT              This sequence was identified as CDM:10210130 by the submitter.

```

For more information on this record e-mail to fly@celera.com.

- * NOTE: This is a 'working draft' sequence.
- * This sequence will be replaced
- * by the finished sequence as soon as it is available and
- * the accession number will be preserved.

FEATURES

source

1. 45605
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 14162 a 9337 c 8941 g 13165 t
ORIGIN

alignment_scores:

Quality: 51.00 Length: 11
Ratio: 4.636 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 90.909

alignment_block:

US-09-471-276-831_COPY_1_16 x AC019542/rev ..

Align seg 1/1 to reverse of: AC019542 from: 1 to: 45605

1 MetSerMetLeuValValPheLeuLeuTrp 11

111:::|||||
37840 ATGGGATGCTTGTGCTTTTATTATTATG 37808

seq_name: gb_pr:HS512B11

seq_documentation_block:

LOCUS HS512B11 64356 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from clone 512B11 on chromosome 6p24-25.
Contains the Desmoplakin I (DPI) gene, ESTs, STS and GSSs.

complete sequence.

ACCESSION AL031058

VERSION AL031058.1 GI:3395507

KEYWORDS HTG; Desmoplakin I; DPI.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 64356)

Phillips, S.

Direct Submission

Submitted (01-OCT-1998) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

requests: clonerequests@sanger.ac.uk

On Aug 5, 1998 this sequence version replaced g1:3355573.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence is the entire insert of clone 512B11. This sequence

has been finished according to sequence map criteria as follows. An

attempt is made to resolve all sequencing problems, such as

compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu). Where the

sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

This sequence was generated from part of bacterial clone contigs of

human chromosome 6, constructed by the Sanger Centre Chromosome 6

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr6

512B11 is from the library RPI3 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong. For further

details see http://hacpac.med.buffalo.edu/ VECTOR: pcTpac2.

FEATURES

source

1. 64356
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"

/map="p24-25"
/clone="Rp3-512B11"
/clone_11b="Rp3-3"
425..504
/note="FLAM_A repeat: matches 48..132 of consensus"
601..899
/note="AluJb repeat: matches 1..301 of consensus"
complement(1093..1568)
/note="match: GSS A0018298"
1606..1722
/note="match: GSS B54321"
complement(1871..2168)
/note="AluJb repeat: matches 301..1 of consensus"
2733..3033
/note="AluJb repeat: matches 1..300 of consensus"
complement(3764..3884)
/note="MIR repeat: matches 238..119 of consensus"
3977..4014
/note="19 copies 2 mer tg 90% conserved"
complement(4480..4775)
/note="AluJg repeat: matches 303..7 of consensus"
complement(5087..5168)
/note="MER46 repeat: matches 234..152 of consensus"
complement(5365..5448)
/note="MER46 repeat: matches 82..1 of consensus"
6625..8578
/note="putative CpG island"
join(6976..7424,21057..21159,23455..23603,24565..24739,
27991..28119,29075..29125,30698..30859,31716..31820,
32693..32788,33120..33245,33776..33928,34525..34679,
35776..35902,36722..36923,37181..37407,39425..39591,
39996..40134,40634..40827,41693..41795,42298..42381,
43118..43225,43803..43901,44614..46908,47981..51913)
/gene="d512B11.1"
/note="match: cDNAs M7830 J05211; match: ESTs AA488478
AA562503 AA857316 A1023316 AA852212 R58704 AA025851 W39488
CO4715 AA099891 AA131128 AA337270 AA340777 AA852213 W74149
AA056401 H91011 T87267 C18467 N57176 AA026480 H83284
AA026654 AA022863 A1022862 AA040016 AA131077 H90899 N29754
W32490 R20904 C03378 H27408 H30146 AA128470 W95408 N84767
AA541923 AA946936 AA644381 AA702424 AA4552300 AA039975
W95070 W19287 A1078449 A1025924 AA088866 AA065627 AA149191
AA515500 AA782066 AA782109 AA602182 T87181 D12062 AA658154
T487933"
/evidence=not-experimental
/product="d512B11.1 (Desmoplakin I (DPI))"
6976..51913
/gene="d512B11.1"
join(7255..7424,21057..21159,23455..23603,24565..24739,
27991..28119,29075..29125,30698..30859,31716..31820,
32693..32788,33120..33245,33776..33928,34525..34679,
35776..35902,36722..36923,37181..37407,39425..39591,
39996..40134,40634..40827,41693..41795,42298..42381,
43118..43225,43803..43901,44614..46908,47981..51217)
/gene="d512B11.1"
/note="match: proteins Q14189 P15924 Q03001 P97395 P97396
P97394"
/codon_start=1
/evidence=not-experimental
/product="d512B11.1 (Desmoplakin I (DPI))"
/protein_id="CAA19927.1"
/db_xref="GI:3702136"
/db_xref="SPTREMBL:O75993"
/translation="MSCNGSHPRINTGLRMIRAESGPDRLREYVSGGCGTSRMYSR
RQVITDONSQGYCQGTMTSRHONONTIOELONCSDCILRALIIVOPELKYDGIQLT
RRELEDEPAONDMETISDLIREBROWGCDYAKOKLLOLODOMRLYVAISPR
VRASSGGGGCTCOSSGMDFTFTVSEGLGMRQAEADMYAEMVAGNDLALONHT
NSHRTGNSIQGYRQDLKTKADLREKTSIYDLREYENYLDAERMRHLQLOQHT
QATSRITMINDCEBELLYDWSDKNTNIAQOEAFSRMSOLEVEKELNKLKQSD
KLVNQHPSADKTEAYMDTLQOWSWIILQITCIVHLKENAAYVQFEEAOSTAYL
QGLQDSIRKQPCDKNMPLQHLLEQIKELEREKILLEYKROVQNLVNSKYKIVOLK
RMDYRSNKPILIRALCDYKQDOKIVHKDEECILKDNNSRKWYVTGPGVDMLVPSV
GLIPPPNPLAVDSLCKIDQYIEYALIALMNQIYINMKSIVSHHYCMIDIEKIRAMTIA

```

KLTMRQEDYKKT1ADLEHLYQER1RNSQSEMRGDDDKRK1QSOFTDOAHNYQTLVT
QUPGYPOHQVTVTTEITTHHCPCODVNNHNV1ETNENENKQSTWMLMELQK1RQ1EHC
BGRMT1KMLPLADQSSHHITV1KINELKSVONDSQDAIEV1NLOK1MLANRGSSEKX
YLONEVFG1PGLQENINGVDGYNS1LCTVRA1LLOAI1QOTD1MK1KYVAR1LEETVC
LDLQVEAYRGC1K1K1ND1NLKKS1L1ATK1PE1LQ1K1Q1H1Q1S1Q1V1P1L1D1L1D1G1C
CEKY1Q1L1D1M1Q1R1D1K1Q1D1P1L1M1D1E1K1Q1K1R1N1R1N1Q1A1F1C1K1L1A1K1R1O1D1E1S
MK1G1D1S1N1T1V1M1R1L1N1E1O1K1N1H1S1E1I1S1O1K1R1K1S1E1V1O1I1A1L1C1A1N1S1I1D1E1Y1Q1A1S1Y1T1S1G1
ET1L1N1P1K1R1M1Q1S1P1S1G1V111Q1E1A1D1H1A1R11E1L1T1R1S1G1D1Y1R1E1S1M1K1S1E1D1K1A1E1L
N1K1I1E1V1E1E1E1L1A1R1A1D1A1N1S1E1N1K1E1F1D1O1N1Q1K1Q1A1F1A1S1E1V1A1L1E1D1K1Q1A1E1L
D1G1K1A1O1N1D1K1C1Y1Q1I1E1L1N1E1K1I1T1R1P1L1E1D1E1K1R1K1S1V1D1R1P1O1O1K1N1D1Y1D1O1L1A1R
O1C1E1N1G1M1O1L1E1S1E1K1A1E1K1E1Y1E1R1L1R11Q1E1G1P1R1E1Y1N1L1A1V1R1N1H1Y1E1M1S
N1L1R1K1Y1E1T1N1T1T1T11K1E1I1S1M1Q1K1H1D1K1N1R1O1D1R11S1R1E1N1D1K1D1E1T1V1N1S1I1O
A1T1E1O1R1R1A1E1N1A1L1O1K1A1C1G1S1E1I1M1Q1K1O1L1E1L1Q1W1O1R1S1E1D1A1R1H1Q1S1E1A1K1T1
Q1D1K1K1E1T1E1R1A1E1F1O1E1A1K1R1M1E1Y1E1N1E1L1S1K1R1N1V1D1E1I1S1L1K1Q1F1E1I1N1T1T1I1H
O1U1T1M1Q1E1D1T1G1Y1A1O1I1D1N1T1R1E1N1S1L1S1E1I1K1R1K1N1T1J1O1T1E1N1R1Y1E1D1I1O1O1K1A1T
G1S1E1Y1O1R1K1O1L1E1V1L1Q1Y1O1T1M1R1E1S1V1Y1T1O1D1A1T1I1D1K1E1I1R1L1O1K1E1T
N1D1R1K1I1E1D1E1N1A1R1O1Y1O1D1O1K1A1S1A1E1T1N1K1I1K1Y1O1E1L1T1R1I1D1E1R1S1O1E1T1V1K
D1O1D1T1R1O1N1S1L1K1E1L1O1K1A1Y1E1E1L1N1R1K1T1A1S1E1D1S1C1K1R1K1L1E1E1B1E1M1R1S1L1E1O1A1I
K1T1N1L1T1O1L1E1Q1A1S1V1K1R1S1E1D1L1Q1O1R1V1D1L1G1H1K1E1K1O1E1L1R1E1L1S1E1V1A1L1R1O1L
L1Q1E1S1V1Q1A1H1R1N1E1H1P1O1K1A1E1D1K1S1R1L1N1E1K1I1E1R1Q1S1T1E1N1T1K1H1L1M1E1E1L1R1
L1R1E1Y1D1L1R1G1R1S1A1D1S1D1K1N1A1T1I1E1L1R1S1Q1I1S1N1R1T1E1L1G1L1I1N1D1O1R1E1N1R1O1E1T
E1K1F1O1K1A1E1A1N1R1O1E1S1K1N1O1C1T1O1Y1O1E1R1S1L1V1K1Y1L1E1O1K1A1R1O1R1L1E1D1L1N1A1K1T
L1E1A1E1T1R1V1O1R1E1C1E1K1O1I1O1N1D1N1O1K1T1O1Y1S1K1E1A1I1K1I1E1S1E1R1E1K1S1E1K1S1E1I1E
R1L1Q1A1E1K1I1E1E1R1C1K1R1K1L1D1S1T1R1E1T1O1S1L1E1T1E1R1S1R1O1R1I1D1L1K1O1P1Y1G1S1H1E1T1O1E1E
W1P1D1T1S1L1V1F1D1G1L1R1K1R1Y1A1M1Q1L1E1C1O1L1D1K1T1L1D1K1L1G1K1S1V1E1V1A1E1I1O1P1L1R1G1A1G
S1I1A1G1A1S1P1K1E1R1Y1S1V1E1A1K1R1K1L1S1P1E1S1T1W1L1E1A1O1A1T1G1I1D1P1H1R1E1K1T1V1S1A1I1
R1D1I1D1F1D1R1O1I1Y1A1E1K1A1I1T1G1D1D1P1E1S1G1Y1S1E1A1I1K1N1L1D1R1E1T1G1R1L1E1A1O1I1S1A1I1
G1V1D1P1V1N1S1V1L1P1K1D1V1A1L1R1G1L1D1R1Y1L1S1L1D1P1R1D1S1O1N1F1D1P1K1K1V1Y1V1O1L1E1R1C
R1E1P1H1T1G1L1L1S1V1O1K1R1S1R1G1O1I1R1P1Y1V1L1N1D1G1L1R1P1T1V1L1E1B1S1G1O1S1Y1D1E1V1C
R1I1K1P1L1O1G1S1S1T1A1G1Y1N1E1T1T1O1K1I1Y1E1A1M1I1C1V1R1P1A1L1E1L1E1A1Q1A1T1F1V1D1P1V1S
N1L1R1L1V1E1A1Y1R1G1L1V1E1F1E1K1L1S1A1E1A1V1Q1V1D1E1S1E1N1I1S1L1F1Q1A1N1K1L1I1K1G1C
I1R1L1E1A1Q1I1A1T1G1I1D1P1K1S1H1R1P1V1D1A1K1R1V1N1E1P1E1S1I1D1S1P1D1K1P1F1D1N1E1T
N1L1Y1L1O1L1E1R1C1I1D1E1T1G1C1L1P1L1K1E1K1Q1V1T1Q1K1L1R1K1R1V1Y1V1P1E1N1K1E1S1V1G
E1A1Y1K1G1L1D1E1T1E1K1E1C1E1O1C1E1M1E1I1T1T1G1S1R1V1Y1L1V1R1K1G1S1O1V1D1O1A1I1D1K1G1L
V1D1R1K1E1P1O1Y1R1G1S1L1T1O1F1A1M1I1S1L1A1N1G1V1S1S1M1G1S1C1V1D1P1S1S1H1E1S1V1S1T1S1
S1S1V1N1U1T1R1S1S1P1D1T1E1S1P1A1I1D1T1E1M1I1D1T1E1K1I1S1T1E1G1R1E1G1V1D1S1T1O1R1L1E1A1O
A1C1T1G1I1H1P1T1T1G1K1L1S1D1A1V1S1Q1V1I1D1D1M1T1R1K1P1A1K1A1I1G1E1G1V1K1R1K1M1S1A1A1E1
V1K1E1W1L1P1E1A1O1R1E1F1Q1Y1L1G1V1D1E1V1H1I1S1T1E1A1I1R1G1F1D1G1R1A1O1M1O1T1S1Y1
A1L1L1T1C1P1T1K1I1S1Y1K1D1A1I1N1S1V1E1D1I1G1L1L1A1S1V1S1G1L1P1S1Y1M1S1A1P1C1S1R1S1G
S1R1S1G1S1R1G1S1R1G1S1R1G1S1F1D1A1T1G1N1S1Y1S1F1S1S1I1G1H1"
8726..8753

repeatL_region
/note:"14 copies 2 mer tt 89% conserved"
9808..9843
/note:"18 copies 2 mer tt 81% conserved"
repeatL_region
complement(11838..11934)
/note:"MIR repeat: matches 146..48 of consensus"
repeatL_region
complement(112765..113075)
/note:"AluJo repeat: matches 302..1 of consensus"
13328..13625
/note:"AluSg repeat: matches 2..300 of consensus"
repeatL_region
complement(113702..113839)
/note:"MIR repeat: matches 192..60 of consensus"
repeatL_region
complement(114492..114786)
/note:"AluSb repeat: matches 297..2 of consensus"
repeatL_region
complement(115382..115693)
/note:"AluJb repeat: matches 295..1 of consensus"
16684..16980
/note:"AluX repeat: matches 1..294 of consensus"
17480..17689
/note:"dJ512B11.1"
/misc_feature
/note:"match: GSS B92850 A0085909"
17617..17920
/note:"AluX repeat: matches 5..301 of consensus"
repeatL_region
complement(118384..118461)
/note:"L1M1C3 repeat: matches 2486..2405 of consensus"
repeatL_region
complement(18476..18598)
/note:"FLAM_C repeat: matches 131..11 of consensus"
19422..19469
/note:"2 copies 24 mer 96% conserved"
repeatL_region
19422..19465
/note:"22 copies 2 mer ac 100% conserved"
repeatL_region
complement(119966..20293)
/note:"AluJo repeat: matches 288..1 of consensus"

```

```

misc_feature 20343..20548
/gene:"dJ512B11.1"
/note:"match: GSS B52084"
repeatL_region 20526..20795
/note:"MER2C repeat: matches 1143..1426 of consensus"
repeatL_region 20782..21014
/note:"MER2C repeat: matches 1323..1538 of consensus"
repeatL_region complement(22529..22700)
/note:"L1M01 repeat: matches 971..805 of consensus"
repeatL_region 22737..23037
/note:"AluX repeat: matches 1..301 of consensus"

alignment_scores:
Quality: 51.00 Length: 12
Ratio: 4.636 Gaps: 0
Percent Similarity: 91.667 Percent Identity: 75.000

alignment_block:
US-09-471-276-831_copy_1_16 x HS512B11 ..
Align seg 1/1 to: HS512B11 from: 1 to: 64356

5 ValValPheLeuLeuEntTgClYValThTtPcLy 16
||||:||||||| |||||||||:||||
29779 GTGATTTTCTATTGGCGCTGCGCTGACTTATGCG 29814

seq_name: gb_hcg:AL590044

seq_documentation_block:
LOCUS AL590044 71990 bp DNA HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 6 clone RP5-875120, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL590044
VERSION AL590044.3 GI:14348513
KEYWORDS HTG: HTGS_PHASE2; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Chromatid: Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 71990)
REFERENCE
AUTHORS Smith,M.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jun 12, 2001 this sequence version replaced gi:13443552.

COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: d1875120
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 71614 bases at least Q40
Consensus quality: 71805 bases at least Q30
Consensus quality: 71883 bases at least Q20
Insert size: 71990; sum-of-contigs
Insert size: 79291; 16.0% error; agarose-fp
Quality coverage: 6.79x in Q20 bases; sum-of-contigs Quality
coverage: 6.16x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
location/Qualifiers
1..71990
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
FEATURES
source

```

```
/chromosome="6"
/clone="RP5-875I20"
/misc_feature 1.:71990
/note="assembly_fragment:01119"
BASE COUNT 20569 a 14394 c 16071 g 20956 t
ORIGIN

alignment_scores:
    quality: 51.00      length: 12
    ratio: 4.636      gaps: 0
    percent similarity: 91.667      percent identity: 75.000

alignment_block:
US-09-471-276-831_COPY_1_16 x AL590044
Align seg 1/1 to: AL590044 from: 1 to: 71990
5 ValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
|||:::||||||| |||||:::|||||
29564 GTGATTTCCTATTGGCGCTGGGCTGACTTATGCG 29599
```


OM of: US-09-471-276-831_COPY_1_16 to: N_Geneseq_1101.* out_format: pfs
Date: Jan 7, 2002 4:54 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framer_pfn model -DEV=xlh  
-O=/cgn2_1/USPTO.spool/US09471276/runat_07012002_142445_19912/app_query.fast_1_274  
-DB=N_Geneseq_1101 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000  
-GAPOP=6.000 -OCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human4.0.cdi -LIST=100 -DOCALLIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USBR=US09471276.@CGN1_1_270 -NCPU=6 -ICPU=3 -LONLOG  
-DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -NO_XLPHY -MNT -THREADS=1
```

Search information block:

Query: US-09-471-276-831_COPY_1_16
Query length: 16
Database: N_Geneseq_1101.*
Database sequences: 930621
Search length: 428662619
Search time (sec): 93.320000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAV88297			87.00	241.05	2.4e+05	276
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ4284			87.00	237.39	3.8e+05	412
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAAC87257			51.00	123.38	85.23	1908
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ01425			51.00	65.89	1.3e+05	1038602
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAAF999			48.00	147.24	3.99	56
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAAF32085			48.00	133.41	23.55	255
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAV88355			48.00	130.81	32.86	339
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAAF32084			48.00	119.26	144.54	1202
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAAF13948			48.00	117.71	176.40	1425
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAV06220			48.00	114.33	271.84	2062
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAV71263			47.50	114.46	267.57	1747
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAAF74241			47.00	135.43	65.54	451
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAAC87258			47.00	114.75	257.72	1453
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAAC00844			46.00	126.37	58.08	300
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAAF93704			46.00	121.02	115.31	539
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH53784			46.00	113.58	299.40	1218
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH17775			46.00	112.72	334.20	1338
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH83568			46.00	111.47	392.48	1535
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH83569			46.00	111.47	392.48	1535
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH14802			46.00	109.21	524.33	1966
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH136160			46.00	109.21	524.33	1966
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH104592			46.00	109.21	524.33	1966
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH17774			46.00	107.42	660.27	2394
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH16647			46.00	104.79	924.24	3191
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH54911			46.00	104.14	1.0e+03	3427
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH54481			46.00	103.08	1.0e+03	3450
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH54639			46.00	103.70	1.1e+03	3599
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAAC76440			46.00	101.61	1.4e+03	4522
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH99053			46.00	100.09	1.7e+03	5345
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ34627			46.00	97.02	2.5e+03	7478
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAV61401			45.00	97.00	3.0e+03	7498
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAI21916			45.00	131.36	30.61	128
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAI47760			45.00	131.36	30.61	128
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH10193			45.00	124.06	78.12	285
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH08124			45.00	124.06	78.12	285
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH12735			45.00	119.45	140.98	472
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH134086			45.00	119.45	140.98	472
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH102644			45.00	119.45	140.98	472

/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH10448	45.00	117.64	177.98	576
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH10038	45.00	116.79	199.40	632
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ4284	45.00	115.25	241.65	748
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAZ4250	45.00	115.25	241.65	748
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH081927	45.00	114.71	259.13	794
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAAC04314	45.00	112.11	361.80	1056
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAAC3416	45.00	109.43	510.00	1416
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH163355	45.00	109.33	516.33	1431
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH163356	45.00	108.69	560.51	1535
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH160820	45.00	108.43	579.36	1579
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH159034	45.00	107.97	614.73	1661
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH15655	45.00	106.52	740.34	1947
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH12905	45.00	106.52	740.34	1947
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH134263	45.00	106.52	740.34	1947
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH163355	45.00	102.78	1.2e+03	2935
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAAC76991	45.00	102.13	1.2e+03	3150
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH163357	45.00	101.93	1.2e+03	3220
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH10225	45.00	89.93	6.2e+03	1139
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH86834	44.50	69.98	7.9e+04	1067
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH86834	44.50	129.16	40.63	140
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH86834	44.00	118.82	152.85	373
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH02781	44.00	118.23	164.91	398
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH30080	44.00	118.19	165.88	400
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH67401	44.00	117.23	187.43	444
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH12554	44.00	117.21	187.93	445
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH86834	44.00	116.81	197.85	465
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH86834	44.00	116.77	198.85	467
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH86834	44.00	116.39	208.85	487
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH83993	44.00	114.16	278.10	622
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH70434	44.00	113.71	294.38	653
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH60963	44.00	112.51	343.49	745
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH93980	44.00	111.70	381.01	814
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH08271	44.00	111.42	395.29	840
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH75486	44.00	111.40	406.32	860
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH33991	44.00	110.46	446.98	933
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH31881	44.00	110.25	458.77	954
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH31640	44.00	110.01	473.44	980
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH80190	44.00	109.72	491.01	1011
/SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAH80194	44.00	108.11	604.17	1207
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH74487	44.00	107.15	682.80	1340
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH61890	44.00	107.12	685.78	1345
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH61890	44.00	106.79	715.71	1395
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH61890	44.00	106.26	745.80	1478
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH61890	44.00	105.64	828.62	1581
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH61890	44.00	105.47	847.66	1612
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH61890	44.00	104.28	987.09	1836
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH61890	44.00	104.27	988.35	1838
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH64765	44.00	103.72	1.1e+03	1953
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH64783	44.00	103.64	1.1e+03	1969
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH64783	44.00	103.54	1.1e+03	1992
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH64783	44.00	101.87	1.3e+03	2392
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH64783	44.00	101.85	1.3e+03	2397
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH64783	44.00	101.85	1.3e+03	2397
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH64783	44.00	101.57	1.4e+03	2471
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH64783	44.00	100.98	1.5e+03	2635
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH64783	44.00	100.45	1.6e+03	2794
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH64783	44.00	100.45	1.6e+03	2794
/SID22/gcgdata/geneseq/geneseq/NA2000.DAT:AAH64783	44.00	100.21	1.7e+03	2868
/SID22/gcgdata/geneseq/geneseq/NA2001.DAT:AAH64783	44.00	99.91	1.7e+03	2964

seq_name: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT:AAV88297

seq_documentation_block:

ID AAV88297 standard: cDNA: 276 BP.

AC AAV88297:

XX 12-FEB-1999 (first entry)

DT EST cDNA GDI177.

DE

XX Expressed sequence tag; secreted protein; haematopoiesis regulator:

KW

```

KW regulation; Identification; ss.
XX
OS Homo sapiens.
XX
PN M09953051-A2.
XX
XX 21-OCT-1999.
XX
XX 09-APR-1999; 99MO-IB00712.
XX
PR 09-APR-1998; 98US-0057719.
PR 28-APR-1998; 98US-0069047.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-038446/03.
DR P-PSDB; AAY64670.
XX
XX Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
XX Claim 1; Page 195; 837pp; English.
XX
XX AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AY65438 represent the EST-related proteins corresponding to AA242265 to
CC AA243052. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used
CC in gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AA242249 to AA24264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
XX
SO Sequence 412 BP: 79 A; 131 C; 119 G; 80 T; 3 other:

Alignment_scores:
Quality: 87.00 Length: 16
Ratio: 5.438 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-471-276-831_COPY_1_16 x AA242284 ..

Align seg 1/1 to: AA242284 from: 1 to: 412

      1 MetSerMetLeuValIlePheLeuLeuLeuTrpClyValThrTrpCly 16
      |||||
      35 AGTCGCATCGTCGTCTTCTCTGTGGTGCGTCACACTGGGCC 82

seq_name: /SIDS2/gcndata/geneseq/geneseqn/AA2001.DAT: AAC87257

seq_documentation_block:
ID AAC87257 standard; cDNA; 1908 BP.
XX
XX AAC87257;
XX
XX 09-MAR-2001 (first entry)
XX

```


US-09-471-276-831_COPY_1_16 x AAZ01425/rev ..
Align seg 1/1 to reverse of: AAZ01425 from: 1 to: 1038602

7 PhaeuleuleuTrpGlyValThrTrpGly 16
|||||
504251 TTCCTCTTTATCGCGGCTTATTTGGGCG 504222

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA07999

seq_documentation_block:

ID AAA07999 standard: DNA; 56 BP.

AC AAA07999;

DT 02-FEB-2001 (first entry)

DE DNA sequence used for marking plants from the Americas.

KW Plant identification: genetic marker: transgenic plant; ds.

OS Unidentified.

PN EP1045037-A1.

PD 18-OCT-2000.

PP 12-APR-1999; 99EP-0106213.

PR 12-APR-1999; 99EP-0106213.

PA (DISC-) DISCOVERY BIOTECH INC.

PA (KAO C.) KAO C.

PI Kao C;

DR WPI: 2000-674087/66.

Genetically marking a plant or plant cell for tracking or identifying plants, especially the distribution of proprietary plants, comprises introducing a DNA sequence having specific information content into a plant genome -
Disclosure: Fig 2; 11pp: English.

The present sequence may be used to identify the geographic origin of a plant. Plants may be genetically marked for identification by introducing a DNA sequence into the plant genome. The DNA sequence is at least 100 base pairs, is unique in the genome, and does not contain a sequence that alters transcription or encodes a functional protein. The sequence is readily detectable and retrievable. The method is useful for tracking or identifying the distribution of proprietary plants.

Sequence 56 BP; 38 A; 12 C; 3 G; 3 T; 0 other;

alignment_scores:
Quality: 48.00 Length: 10
Ratio: 5.333 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 70.000

alignment_block:

US-09-471-276-831_COPY_1_16 x AAA07999/rev ..

Align seg 1/1 to reverse of: AAA07999 from: 1 to: 56

7 PhaeuleuleuTrpGlyValThrTrpGly 16
|||||
54 TTTTCTTTACTTTGGGCTTACTTTGGGCT 25

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF32085

seq_documentation_block:
ID AAF32085 standard: cDNA; 255 BP.

AC AAF32085;

DT 12-APR-2001 (first entry)

DE Human M019 open reading frame.

KW Human; M010; M019; M024; cancer; Obesity; diabetes; osteoporosis; asthma; nervous system disorder; pain; stroke; hypercholesterolaemia; chronic obstructive pulmonary disease; hyperlipidaemia; atherosclerosis; hyperlipoproteinemia; arteriosclerosis; coronary artery disease; ss.

OS Homo sapiens.

PN WO200100644-A1.

PD 04-JAN-2001.

PP 30-JUN-2000; 2000WO-US18097.

PR 30-JUN-1999; 99US-0345293.

PR 30-JUN-1999; 99US-0345680.

PA (MILL-) MILLENNIUM PHARM INC.

PI Khodadoust M;

DR WPI: 2001-112434/12.

DR P-PSDB; AAB66893, AAB66894, AAB66895.

New gene encoding proteins designated M010, M019, M024, useful for the diagnosis, prophylaxis and treatment of cancer, obesity, diabetes, osteoporosis, asthma, atherosclerosis and pulmonary disorders -
Claim 1; Fig 2; 98pp: English.

The present invention relates to novel human coding sequences and proteins designated M010 (AAF32082 and AAB66884), M019 (AAF32084 and AAB66893) and M024 (AAF32086 and AAB66896). M010, M019 and M024 proteins and compounds which modulate their activity or which bind M010, M019 and/or M024 are useful for prevention and treatment of a variety of disorders including cancer, obesity, diabetes, osteoporosis, asthma, central and peripheral nervous system disorders, pain, stroke, chronic obstructive pulmonary disease, hypercholesterolaemia, hyperlipidaemia, hyperlipoproteinemia, atherosclerosis, arteriosclerosis and coronary artery disease. The present sequence is the open reading frame for human M019.

Sequence 255 BP; 68 A; 41 C; 78 G; 68 T; 0 other;

alignment_scores:
Quality: 48.00 Length: 11
Ratio: 5.333 Gaps: 0
Percent Similarity: 81.818 Percent Identity: 72.727

alignment_block:

US-09-471-276-831_COPY_1_16 x AAF32085 ..

Align seg 1/1 to: AAF32085 from: 1 to: 255

6 ValPhaeuleuleuTrpGlyValThrTrpGly 16
|||||
34 GTCCTTTTACTCTCTGCGGCTTATCTGCGGA 66

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT:AAV88355

seq_documentation_block:

ID AAV88355 standard: cDNA; 339 BP.

AC AAV88355;

```
XX 12-FEB-1999 (first entry)
DE EST clone GP56.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; actinin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
XX Homo sapiens.
XX
XX M09845437-A2.
XX
XX 15-OCT-1998.
XX
XX 10-APR-1998; 98WO-US06956.
XX
XX 10-APR-1997; 97US-0837312.
XX
XX (GENY ) GENETICS INST INC.
XX
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Werberg D;
XX Racle LA, Spaulding V, Treacy M;
XX
XX MPI: 1999-070078/06.
XX
XX New polynucleotides encoding human secreted proteins - derived from
XX e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX ovary, pituitary, retina and colon cDNA libraries
XX
XX Claim 1: Page 368; 641pp: English.
XX
XX The present sequence represents an expressed sequence tag (EST), and is
XX a polynucleotide of the invention. The polynucleotides of the invention
XX are all secreted EST sequences isolated from a variety of human tissue
XX sources. The EST sequences and proteins encoded by them are predicted to
XX have useful biological activities which would make them suitable for
XX treating, preventing or ameliorating medical conditions in humans and
XX animals, although no supporting data is given. Suggested activities
XX include nutritional activity, immune stimulating or suppressing activity,
XX haematopoiesis regulating activity, tissue growth activity,
XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, cadherin/tumour invasion suppressor activity, tumour inhibition
XX activity. The EST sequences are also stated to be useful for gene
XX therapy.
XX
XX Sequence 339 BP; 90 A; 74 C; 47 G; 128 T; 0 other:
XX
XX
XX alignment_scores:
XX      Quality: 48.00      Length: 14
XX      Ratio: 4.364      Gaps: 0
XX      Percent Similarity: 78.571      Percent Identity: 50.000
XX
XX alignment_block:
XX US-09-471-276-831_COPY_1_16 x AAV88355/rev ..
XX
XX Align seg 1/1 to reverse of: AAV88355 from: 1 to: 339
XX
XX 2 SerMetLeuValAlaPheLeuLeuTrpGlyValThrTrp 15
XX |||::: ::|||::: |||::: |||
XX 74 AGTATATAAAATGCTTTATCACACTCTGGGACCTGTGG 33
XX
XX seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF32084
XX
XX seq_documentation_block:
XX ID AAF32084 standard; cDNA: 1202 BP.
XX
XX AC AAF32084;
XX
XX DT 12-APR-2001 (first entry)
```

```
XX
DE Human M019 coding sequence.
XX
XX Human; M010; M019; M024; cancer; obesity; diabetes; osteoporosis; asthma;
XX nervous system disorder; pain; stroke; hypercholesterolemia;
XX chronic obstructive pulmonary disease; hyperlipidemia; atherosclerosis;
XX hyperlipoproteinemia; arteriosclerosis; coronary artery disease; ss.
XX
XX Homo sapiens.
XX
XX M0200100644-A1.
XX
XX 04-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US18097.
XX
XX 30-JUN-1999; 99US-0345293.
XX
XX 30-JUN-1999; 99US-0345680.
XX
XX (MILL-) MTLENNIUM PHARM INC.
XX
XX Khodadoust M;
XX
XX MPI: 2001-112434/12.
XX
XX P-PsDB: AAB66893, AAB66894, AAB66895.
XX
XX New gene encoding proteins designated M010, M019, M024, useful for the
XX diagnosis, prophylaxis and treatment of cancer, obesity, diabetes,
XX osteoporosis, asthma, atherosclerosis and pulmonary disorders -
XX
XX Claim 1: Flg 2; 98pp: English.
XX
XX The present invention relates to novel human coding sequences
XX and proteins designated M010 (AAF32082 and AAB66894), M019 (AAF32084 and
XX AAB66893) and M024 (AAF32086 and AAB66896). M010, M019 and M024 proteins
XX and compounds which modulate their activity or which bind M010, M019
XX and/or M024 are useful for prevention and treatment of a variety of
XX disorders including cancer, obesity, diabetes, osteoporosis, asthma,
XX central and peripheral nervous system disorders, pain, stroke, chronic
XX obstructive pulmonary disease, hypercholesterolemia, hyperlipidaemia,
XX hyperlipoproteinemia, atherosclerosis, arteriosclerosis and coronary
XX artery disease. The present sequence is the coding sequence for human
XX M019.
XX
XX Sequence 1202 BP; 406 A; 226 C; 309 G; 261 T; 0 other:
XX
XX
XX alignment_scores:
XX      Quality: 48.00      Length: 11
XX      Ratio: 5.333      Gaps: 0
XX      Percent Similarity: 81.818      Percent Identity: 72.727
XX
XX alignment_block:
XX US-09-471-276-831_COPY_1_16 x AAF32084 ..
XX
XX Align seg 1/1 to: AAF32084 from: 1 to: 1202
XX
XX 6 ValPheLeuLeuLeuTrpGlyValThrTrpGly 16
XX |||::: |||::: |||::: |||
XX 364 GTCCTTTACTCTCTGGGCTTATCTGCGGA 396
XX
XX seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1998.DAT:AAAX13948
XX
XX seq_documentation_block:
XX ID AAAX13948 standard; DNA: 1425 BP.
XX
XX AC AAAX13948;
XX
XX DT 31-MAR-1999 (first entry)
XX
XX DE H. pylori GHPO 1365 gene.
XX
XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
```

XX	peptic ulcer disease; ss.
OS	XX
XX	Helicobacter pylori.
XX	
FH	Key
FT	Location/Qualifiers
FM	97..1374
XX	/*tag" a
XX	
PN	MO9843478-A1.
PD	XX
XX	08-OCT-1998.
XX	
PE	01-APR-1998; 98MO-US06371.
XX	
PR	29-JUL-1997; 97US-0902615.
PR	01-APR-1997; 97US-0833457.
PR	24-JUN-1997; 97US-0881227.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PA	(INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI	Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX	
DR	WPI: 1998-542293/46.
DR	P-PSDB; AAM98229.
XX	
PT	New Isolated Helicobacter polynucleotides - used to develop products
PT	for the diagnosis, prevention and treatment of Helicobacter
PT	infections and gastrointestinal diseases
XX	
PS	Claim 1; Page 166-168; 2054pp; English.
CC	
CC	This sequence represents a polynucleotide of the invention. It was
CC	isolated from Helicobacter pylori and encodes a H-pylori GPO protein.
CC	The polypeptides can be used for preventing or treating Helicobacter
CC	infections, and gastroduodenal diseases associated with these
CC	infections, including acute, chronic, and atrophic gastritis, and peptic
CC	ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC	for the production of antibodies. The products can also be used for
CC	detection and diagnosis.
XX	
SO	Sequence 1425 BP; 415 A; 275 C; 232 G; 503 T; 0 other;
	Alignment_scores:
	Quality: 48.00 Length: 13
	Ratio: 5.333 Gaps: 0
	Percent Similarity: 69.231 Percent Identity: 61.538
	Alignment_block:
	US-09-471-276-831_COPY_1_16 x AAX13948 ..
	Align seg 1/1 to: AAX13948 from: 1 to: 1425
	4 LeuValValPheLeuLeuLeuLeuTrpGlyValThrTrpIly 16
	::::
	266 CTAGAAATAGTCCTTTCCTTTCGCGGCTTTTCGCGCG 304
	seq_name: /SID02/gcgdata/geneseq/geneseqn/AA199.DAT:AAZ06220
	seq_documentation_block:
	ID AAZ06220 standard: DNA: 2062 BP.
XX	
AC	AAZ06220;
XX	
DT	30-SEP-1999 (first entry)
XX	
DE	Human secreted protein gene No. 2.
XX	
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;

```

KW inflammation; ischaemic shock; Alzheimer's disease; retinosis; AIDS;
KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
XX OS Homo sapiens.
XX
XX PN W09335158-AL.
XX
XX PD 15-JUL-1999.
XX
XX PF 06-JAN-1999; 99WO-US00108.
XX
XX PR 07-JAN-1998; 98US-0070704.
XX PR 07-JAN-1998; 98US-0070657.
XX PR 07-JAN-1998; 98US-0070658.
XX PR 07-JAN-1998; 98US-0070692.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Brewer LA, Duan RD, Edner R, Tafleur DM, Ni J;
XX PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
XX
XX DR WPI: 1999-444190/37.
XX DR P-PSDB: AA38387.
XX
XX PS New isolated human genes and the secreted polypeptides they encode
XX PS Claim 1; Page 150; 227pp; English.
XX
XX PS This sequence represents a nucleic acid molecule which encodes a
XX CC secreted human protein. The gene number is given in the descriptor line.
XX CC The gene can be used to generate fusion proteins by linking to the gene
XX CC to a human immunoglobulin Fc portion (e.g. AA06210) for increasing the
XX CC stability of the fused protein as compared to the human protein only.
XX CC The invention relates to 36 novel genes and their fragments (nucleic
XX CC acid sequences: AA06219-206263; amino acid sequences AA38386-38436)
XX CC which are useful for preventing, treating or ameliorating medical
XX CC conditions e.g. by protein or gene therapy. Also, pathological
XX CC conditions can be diagnosed by determining the amount of the new
XX CC polypeptides in a sample or by determining the presence of mutations in
XX CC the new polynucleotides. Specific uses are described for each of the 36
XX CC polynucleotides, based on which classes they are most highly expressed in
XX CC (see AA06219 for described uses).
XX
XX SQ Sequence 2062 BP; 666 A; 265 C; 335 G; 788 T; 8 other;

alignment_scores:
      Quality: 48.00      Length: 13
      Ratio: 4.800      Gaps: 0
Percent Similarity: 76.923      Percent Identity: 61.538

alignment_block:
US-09-471-276-831_COPY_1_16 x AA06220      ..
Align seg 1/1 to: AA06220 from: 1 to: 2062
      3 MetLeuValValPheLeuLeuLeuTrpClyValThrTrp 15
      |||::: |||:::||||||| |||||
1854 ATGTGTTTGAATTTGTACTATGTCGGGGATGACTTGG 1892

seq_name: /SID52/gcdata/geneseq/geneseqn/NA199.DAT:AAV71763
seq_documentation_block:
ID AAV71763 standard; DNA; 1747 BP.
XX
XX AC AAV71763;
XX
XX DT 09-MAR-1999 (first entry)
XX
XX K. pneumoniae DNA probe KP-85-43.
XX

```

```

KM Detection; diagnosis; infection; KP-77-46; KP-85-43; KP-98-22;
KW KP-98-33; KP-110-32; probe; ds.
XX
XX Klebsiella pneumoniae.
XX
XX MO9842843-A1.
XX
XX PD 01-OCT-1998.
XX
XX PF 23-MAR-1998; 98MO-JP01286.
XX
XX PR 25-MAR-1997; 97JP-0071082.
XX
XX PA (FUSO ) FUSO PHARM IND LTD.
XX
XX PI Abe K, Keshi H, Matsuhisa A, Ueyama H;
XX WPI; 1999-045140/04.
XX
XX PT DNA probes for diagnosis of Klebsiella pneumoniae infections - are
XX obtained by HindIII digestion of genomic K. pneumoniae DNA and
XX selection of bacterial-binding fragments
XX
XX PS Claim 2; Page 17-18; 32pp; Japanese.
XX
XX CC This sequence is a DNA probe used for the diagnosis of infection with
XX Klebsiella pneumoniae. The detection method involves isolating genomic
XX DNA from this bacterium, digesting it with HindIII restriction enzyme,
XX CC and identifying clones capable of binding to the bacterium then isolating
XX CC and sequencing the HindIII fragments involved. The probes have sequences
XX KP-77-46, KP-85-43, KP-98-22, KP-98-33 and KP-110-32. The probes allow
XX CC for simple and highly specific detection of Klebsiella pneumoniae in
XX biological samples such as blood.
XX
XX SQ Sequence 1747 BP; 534 A; 367 C; 403 G; 443 T; 0 other;
XX
XX Alignment_scores:
XX Quality: 47.50 Length: 18
XX Ratio: 3.167 Gaps: 1
XX Percent Similarity: 83.333 Percent Identity: 44.444
XX
XX Alignment_block:
XX US-09-471-276-831_COPY_1_16 x AAV71763/rev ..
XX
XX Align seg 1/1 to reverse of: AAV71763 from: 1 to: 1747
XX
XX 1 MetSerMetLeuValValPheLeuLeu.....TrrGlyValIth 14
XX :::::::::::::::::::::::::::::::
XX 1442 GAGCAAAATGATGATATTATTCCTGCTGTCGATGCAATGACCTGGGAAACAC 1393
XX 14 Trrp 15
XX ||||
XX 1392 ATGG 1389
XX
XX seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA74241
XX
XX seq_documentation_block:
XX ID AAA74241 standard; DNA: 451 BP.
XX
XX AC AAA74241;
XX
XX CC 29-NOV-2000 (first entry)
XX
XX DE Lobliolly pine SSR locus R1PPT165.
XX
XX Lobliolly pine; Simple Sequence Repeat; SSR; microsatellite DNA repeat;
XX genetic marker; mapping; inheritance study; population genetics study;
XX plant breeding programme; ss.
XX
XX OS Pinus taeda.
XX
XX PN MO200042210-A2.

```

```

PD 20-JUL-2000.
XX
XX 06-JAN-2000: 2000MO-US00325.
XX
XX 15-JAN-1999: 99US-0232884.
XX
XX 19-JAN-1999: 99US-0232785.
XX
XX (INTO ) INT PAPER CO.
PA (ECHT/) ECHT C. S.
PA (NELS/) NELSON C. D.
PA (USDA ) US SEC OF AGRIC.
XX
PI Echt CS, Nelson CD;
XX
XX WPI: 2000-482836/42.
XX
XX Polynucleotide having simple sequence repeat useful as markers in
PT plants for genetic characterization e.g. genetic mapping study, an
PT inheritance study of a commercially important trait in a plant breeding
PT program -
XX
XX Claim 1: Pages 32-33; 57pp; English.
XX
XX The present sequence is a loblolly pine Simple Sequence Repeat (SSR)
CC locus. SSRs are also known as microsatellite DNA repeats. The present
CC sequence is useful as a genetic marker for genetic mapping, population
CC genetics studies and inheritance studies in various plant breeding
CC programmes.
XX
XX Sequence 451 BP: 209 A: 57 C: 63 G: 122 T: 0 other:
XX
XX
XX alignment_scores:
XX      Quality: 47.00      Length: 13
XX      Ratio: 3.615      Gaps: 0
XX      Percent Similarity: 100.000      Percent Identity: 53.846
XX
XX alignment_block:
XX US-09-471-276-831_COPY_1_16 x AAA74241/rev ..
XX
XX Align seg 1/1 to reverse of: AAA74241 from: 1 to: 451
XX
XX      3 MetLeuVal1PheLeuLeuLeuTrpGlyVal1nhrTrp 15
XX      :::::::::::::::::::::::::::::::
XX      165 TTATTATTGTTGTTGTTGTTGTTATGGGGCATTTACATAC 127
XX
XX seq_name: /SIS2/9/cgdata/geneseq/geneseqn/NA2001.DAT:AAc87258
XX
XX seq_documentation_block:
XX ID AAC87258 standard: cDNA: 1453 BP.
XX
XX AAC87258:
XX
XX 09-MAR-2001 (first entry)
XX
XX DT
XX
XX Mouse liver growth hormone-induced cDNA clone 45, SEQ ID NO:9.
XX
XX
XX Mouse: growth hormone; GH regulatable gene; liver pathology; hypertrophy;
KW hepatocellular lesion; hyperplasia; altered expression level; clone 45;
KW diagnostic marker; gigantism; acromegaly; diabetes; hepatotropic;
KW transgenic animal; drug screening; drug discovery; mouse; murine; ss.
XX
XX Mus sp.
XX
XX OS
XX
XX PN WO20006787-A2.
XX
XX
XX 09-NOV-2000.
XX
XX PD
XX
XX 05-MAY-2000: 2000MO-US12366.
XX
XX PF
XX
XX 05-MAY-1999: 99US-0132663.
XX
XX

```

```
PA (UYOH-) UNIV OHIO.
XX
PI Korchick JJ, Tlong J;
XX
DR WPI: 2001-007239/01.
XX
DR P-PSDR: AAB48740, AAB48741, AAB48742, AAB48743, AAB48744, AAB48745,
XX AAB48746, AAB48747, AAB48748.
XX
PT Diagnosing abnormal levels of growth hormone activity in liver
PT comparing assaying growth transcriptional activity and protein
PT expression level of hormone-regulatable liver genes, as diagnostic
PT markers of liver pathology
XX
XX
PS Claim 1; Page 55-56; 65pp; English.
XX
CC The invention relates to a method of diagnosing abnormal levels of
CC growth hormone (GH) activity in the liver, or predicting a change in the
CC condition of the liver in response to abnormal GH activity. The method
CC involves correlating the level of expression of certain specific genes
CC with the level of GH activity in the liver, or with an expected change
CC in the condition of the liver as the result of GH activity. Excessive GH
CC activity in the liver is thought to be deleterious to health, causing an
CC increase in liver size as a consequence of both hyperplasia and
CC hepatocyte hypertrophy, and hepatocellular lesions which progress with
CC age. Studies in transgenic mice which express high levels of bovine
CC growth hormone identified a number of genes whose expression in the
CC liver is altered by high GH levels. The genes which are upregulated are
CC those encoding alpha-fetoprotein, corticosteroid binding globulin,
CC fetuin, rab8-interacting protein, paraoxonase-3, cytochrome P45011A,
CC S-2 hydroxyacid oxidase, interferon alpha/beta receptor, growth hormone
CC receptor, procasome z-subunit, and coagulation factor V. Two novel
CC genes, clone 5 (AAC87257) and clone 45 (AAC87258) are also upregulated
CC in response to abnormally high GH levels. Conversely, expression of the
CC gene encoding 3-beta-hydroxysteroid dehydrogenase/delta-5-delta-4
CC isomerase appears to be downregulated. The invention also relates to
CC transgenic nonhuman mammals comprising a GH-induced transgene which
CC exhibit or have a propensity to develop a liver pathology; an assay for
CC drugs which inhibit the development of, or which treat a liver pathology,
CC comprising administering the drug to the transgenic animal; and
CC preventing or treating a liver pathology in a patient comprising
CC administering a drug which inhibits the expression of a GH-induced gene.
CC The method of the invention is used for diagnosing abnormal levels of GH
CC activity in the liver or predicting a change in the condition of the
CC liver in response to abnormal levels of GH activity. The GH-regulatable
CC liver genes and proteins are useful as diagnostic markers of liver
CC pathology. Assays for the expression of these genes is useful for the
CC diagnosis of liver pathologies associated with gigantism or acromegaly or
CC with diabetes, as other causative agents may act directly or indirectly
CC upon the same genes. The present sequence represents the novel mouse
CC liver cDNA clone 45, expression of which is upregulated by abnormal GH
CC levels.
XX
XX
SO Sequence 1453 BP; 363 A; 392 C; 352 G; 346 T; 0 other;

alignment_scores:
    Quality: 47.00      Length: 14
    Ratio: 3.917        Gaps: 0
    Percent Similarity: 85.714    Percent Identity: 64.286

alignment_block:
US-09-471-276-831_COPY_1_16 x AAC87258
..
Align seg 1/1 to: AAC87258 from: 1 to: 1453

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThr 14
|||||:|||||:|||||:|||||:|||||
20 ATGCTCTCTGCTGCTACTGCTGCTGCTGCTGCTGCTTCTACT 61

seq_name: /sids2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA00844
seq_documentation_block:
ID AAA00844 standard; cDNA; 300 BP.
```

```
XX
AC AAA00844;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:835.
XX
KM Human: colon cancer; tumour; diagnosis; gene expression product;
KM probe; detection; cancerous state; metastasis; identification;
KM breast cancer; oestrogen receptor-positive breast cancer; therapy;
KM oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN M09958675-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99MO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEO INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
DR WPI: 2000-126369/11.
XX
PT Polynucleotide library used to determine cancerous states of mammalian
PT cells -
XX
PS Claim 1; Page 384; 1097pp; English.
XX
CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present
CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotide sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
XX
SO Sequence 300 BP; 72 A; 46 C; 96 G; 86 T; 0 other;

alignment_scores:
    Quality: 46.00      Length: 15
    Ratio: 3.833        Gaps: 0
    Percent Similarity: 80.000    Percent Identity: 46.667

alignment_block:
US-09-471-276-831_COPY_1_16 x AAA00844
..
Align seg 1/1 to: AAA00844 from: 1 to: 300

2 SerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
```

|||||
78 AGCGTGTGCTGCTTACTGATCTGTGGCGTGGCAATGGGGA 122

seq_name: /SID52/gcdata/geneseq/geneseqn/NA1999.DAT:AAV89754

seq_documentation_block:

ID AAV89754 standard: cDNA: 361 BP.

AC AAV89754;

DT 15-FEB-1999 (first entry)

DE EST clone CS485.

Human; secreted protein; expressed sequence tag; EST; haematopoiesis; tissue growth; activin; inhibitor; chemotaxis; chemokinesis; haemostatic; receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour; gene therapy; ss.

OS Homo sapiens.

PN WO9845436-A2.

PD 15-OCT-1998.

PF 10-APR-1998; 98WO-US06955.

PR 10-APR-1997; 97US-0838821.

PA (GENY) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PI Racle LA, Spaulding V, Treacy M;

DR WPI: 1999-070077/06.

New polynucleotides encoding human secreted proteins - derived from e.g. human blood, kidney, foetal lung, placenta, testes, brain, ovary, pituitary, retina and colon cDNA libraries.

PS Claim 1: Page 319; 618pp: English.

CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.

SO Sequence 361 BP; 88 A; 112 C; 80 G; 81 T; 0 other;

alignment_scores:

Quality: 46.00 Length: 22
Ratio: 3.067 Gaps: 1
Percent Similarity: 68.182 Percent Identity: 36.364

alignment_block:

US-09-471-276-831_COPY_1_16 x AAV89754/rev ..

Align seg 1/1 to reverse of: AAV89754 from: 1 to: 361

1 MeserleuValValPheLeuLeuLeuTrrpGlyValThrTrp 11
|||||
256 ATGAGTTTGATGTGACCTGAGAGCTGAACCTGTGATTCATGATGTTCTG 207
11 pGlyValThrTrpGly 16

|||||
206 GAGCGTGACCTGGCGGT 191

seq_name: /SID52/gcdata/geneseq/geneseqn/NA2001.DAT:AAF93704

seq_documentation_block:

ID AAF93704 standard: cDNA: 539 BP.

AC AAF93704;

DT 21-MAY-2001 (first entry)

DE cDNA encoding SRT protein isolated from testis tissue SEQ ID 525.

Human; SRT; gene therapy; gene mapping; tissue typing; ss.

OS Homo sapiens.

PN WO200107611-A2.

PD 01-FEB-2001.

PF 21-JUL-2000; 2000WO-US20006.

PR 26-JUL-1999; 99US-0145701.

PA (GETH) GENENTECH INC.

PI Baker KP, Goddard A, Wood WT;

DR WPI: 2001-112729/12.

New isolated nucleic acid molecule encoding a SRT polypeptide is useful for production of recombinant SRT polypeptides, gene mapping, diagnosing genetic disorders and for gene therapy -

PS Claim 2: Fig 525; 663pp: English.

CC Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding
CC human SRT proteins. The cDNA sequences are isolated from various
CC different human tissue cDNA libraries. The invention relates to a method
CC for detecting cDNA encoding an SRT protein, a vector containing cDNA
CC encoding SRT, a host cell transformed with the vector, an isolated SRT
CC polypeptide, and an antibody which binds to SRT. The polynucleotide
CC sequence can be used in gene therapy and is useful in the recombinant
CC production of SRT polypeptides, as a hybridisation probe to screen
CC libraries to isolate cDNAs with sequence identity to SRT polypeptides, to
CC map the gene encoding the SRT polypeptides and analysing genetic
CC disorders, tissue typing and disease tissue detection. The SRT
CC polynucleotide sequences can be used in polymerase chain reaction,
CC screening for new therapeutic molecules and generation of antisense RNA
CC and DNA.

SO Sequence 539 BP; 159 A; 107 C; 135 G; 137 T; 1 other;

alignment_scores:

Quality: 46.00 Length: 14
Ratio: 4.182 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 50.000

alignment_block:

US-09-471-276-831_COPY_1_16 x AAF93704/rev ..

Align seg 1/1 to reverse of: AAF93704 from: 1 to: 539

2 MeserleuValValPheLeuLeuLeuTrrpGlyValThrTrp 15
|||||
383 AGCACAACCTTTATTATGCGCTTCTCTGGGAGATCACTTGG 342
seq_name: /SID52/gcdata/geneseq/geneseqn/NA2001.DAT:AAH53784
seq_documentation_block:


```
alignment_block:
US-09-471-276-831_COPY_1_16 x AAX17775  ..
Align seg 1/1 to: AAX17775 from: 1 to: 1338
      2 SerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrp 15
      ||| |||::: ||| ||| |||::: ||| |||
      845 TCCCGCGCTCTGATGACCTGCCACTCTGGGCGCTGACCTGG 886

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT: AAX83568

seq_documentation_block:
ID AAX83568 standard; DNA: 1535 BP.
XX
AC AAX83568;
XX
DT 21-DEC-1999 (first entry)
XX
DE 16S rDNA gene fragment from marine bacterium, isolate K2-1.
XX
KW Monitoring: oil; contamination: sea water; detection: flagellum;
KW Gram-negative bacterium; Proteobacteria; glucose; carbon source;
KW alkane; 16S rDNA gene; ds.
XX
OS Proteobacteria.
XX
PN JP11243967-A.
XX
PD 14-SEP-1999.
XX
PE 04-MAR-1998; 98JP-0069399.
XX
PR 04-MAR-1998; 98JP-0069399.
XX
PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
XX
DR WPI: 1999-564435/48.
XX
XX
PT Monitoring of oil contamination of sea water - where oil contamination
PT is evaluated by detection of a microbe having properties from e.g.
PT having no flagellum, being a Gram-negative bacterium, belonging to
PT Proteobacteria, gamma subdivision, etc.2
XX
PS Claim 3; Page 7; 15pp; Japanese.
XX
CC The invention relates to a method for monitoring oil contamination of
CC sea water by detecting, in the sea water, a microbe having the following
CC properties: (1) it has no flagellum; (2) it is a Gram-negative bacterium;
CC (3) it belongs to Proteobacteria, gamma subdivision; (4) it cannot
CC assimilate glucose as a single carbon source; and (5) it efficiently
CC assimilates at least one of 10-30C n-alkanes. This sequence represents
CC a fragment of the 16S rDNA gene from the microbe of the invention,
CC isolate K2-1.
XX
SQ Sequence 1535 BP; 376 A; 349 C; 494 G; 316 T; 0 other:

alignment_scores:
      Quality: 46.00      Length: 7
      Ratio: 6.571      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 85.714

alignment_block:
US-09-471-276-831_COPY_1_16 x AAX83568  ..
Align seg 1/1 to: AAX83568 from: 1 to: 1535
      10 LeuTrpGlyValThrTrpGly 16
      ||| ||| |||::: ||| ||| |||
      132 TTCTGGCGGATTAACCTTGGGGA 152

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT: AAX83569
```

```
seq_documentation_block:
ID AAX83569 standard; DNA: 1535 BP.
XX
AC AAX83569;
XX
DT 21-DEC-1999 (first entry)
XX
DE 16S rDNA gene fragment from marine bacterium, isolate K3-3.
XX
KW Monitoring: oil; contamination: sea water; detection: flagellum;
KW Gram-negative bacterium; Proteobacteria; glucose; carbon source;
KW alkane; 16S rDNA gene; ds.
XX
OS Proteobacteria.
XX
PN JP11243967-A.
XX
PD 14-SEP-1999.
XX
PE 04-MAR-1998; 98JP-0069399.
XX
PR 04-MAR-1998; 98JP-0069399.
XX
PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
XX
DR WPI: 1999-564435/48.
XX
XX
PT Monitoring of oil contamination of sea water - where oil contamination
PT is evaluated by detection of a microbe having properties from e.g.
PT having no flagellum, being a Gram-negative bacterium, belonging to
PT Proteobacteria, gamma subdivision, etc.2
XX
PS Claim 3; Page 7-8; 15pp; Japanese.
XX
CC The invention relates to a method for monitoring oil contamination of
CC sea water by detecting, in the sea water, a microbe having the following
CC properties: (1) it has no flagellum; (2) it is a Gram-negative bacterium;
CC (3) it belongs to Proteobacteria, gamma subdivision; (4) it cannot
CC assimilate glucose as a single carbon source; and (5) it efficiently
CC assimilates at least one of 10-30C n-alkanes. This sequence represents
CC a fragment of the 16S rDNA gene from the microbe of the invention,
CC isolate K3-3.
XX
SQ Sequence 1535 BP; 375 A; 350 C; 496 G; 314 T; 0 other:

alignment_scores:
      Quality: 46.00      Length: 7
      Ratio: 6.571      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 85.714

alignment_block:
US-09-471-276-831_COPY_1_16 x AAX83569  ..
Align seg 1/1 to: AAX83569 from: 1 to: 1535
      10 LeuTrpGlyValThrTrpGly 16
      ||| ||| |||::: ||| ||| |||
      132 TTCTGGCGGATTAACCTTGGGGA 152

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT: AAI14802

seq_documentation_block:
ID AAI14802 standard; DNA: 1966 BP.
XX
AC AAI14802;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #4735 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
```

```
KW cervical cancer; ss.
XX
XX Homo sapiens.
OS
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID NO 4735; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENP). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification. The sequence was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1966 BP; 406 A; 402 C; 610 G; 548 T; 0 other:

alignment_scores:
    Quality: 46.00      Length: 22
    Ratio: 3.067        Gaps: 1
    Percent Similarity: 68.182   Percent Identity: 36.364

alignment_block:
US-09-471-276-831_COPY_1_16 x AA114802 ..
Align seg 1/1 to: AA114802 from: 1 to: 1966

1 MetSerMetLeuValVal.....PheLeuLeuLeuTr 11
|||||:|||||:|||||:|||||:|||||:|||||:
1282 ATGAGTTTGTGATGCTGACCTGGAGCTGAACCTGTGACTTCATGATGTTCTG 1331
11 pGIyValtThrTPGly 16
|:|||||:|||||:
1332 GAGCCTGACCTGGGCT 1347

seq_name: /SID52/gcdata/geneseq/geneseqn/NA2001.DAT:AA136160
seq_documentation_block:
ID AA136160 standard; DNA; 1966 BP.
XX
XX AA136160;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #4846 used to measure gene expression in human placenta sample.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
```

```
KW genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID NO 4846; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 1966 BP; 406 A; 402 C; 610 G; 548 T; 0 other:

alignment_scores:
    Quality: 46.00      Length: 22
    Ratio: 3.067        Gaps: 1
    Percent Similarity: 68.182   Percent Identity: 36.364

alignment_block:
US-09-471-276-831_COPY_1_16 x AA136160 ..
Align seg 1/1 to: AA136160 from: 1 to: 1966

1 MetSerMetLeuValVal.....PheLeuLeuLeuTr 11
|||||:|||||:|||||:|||||:|||||:|||||:
1282 ATGAGTTTGTGATGCTGACCTGGAGCTGAACCTGTGACTTCATGATGTTCTG 1331
11 pGIyValtThrTPGly 16
|:|||||:|||||:
1332 GAGCCTGACCTGGGCT 1347

seq_name: /SID52/gcdata/geneseq/geneseqn/NA2001.DAT:AA104592
seq_documentation_block:
ID AA104592 standard; DNA; 1966 BP.
XX
XX AA104592;
XX
XX 09-OCT-2001 (first entry)
XX
XX Probe #4583 used to measure gene expression in human breast sample.
XX
XX Probe; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX WO200157270-A2.
```

```
XX 09-AUG-2001.
PD
XX 29-JAN-2001; 2001WO-US000661.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
PT
XX
PS Claim 25; SEQ ID NO 4583; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridizes at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1966 BP; 406 A; 402 C; 610 G; 548 T; 0 other;

alignment_scores:
    Quality: 46.00      Length: 22
    Ratio: 3.067        Gaps: 1
    Percent Similarity: 68.182    Percent Identity: 36.364

alignment_block:
US-09-471-276-831_COPY_1_16 x AA104592 ..
Align seg 1/1 to: AA104592 from: 1 to: 1966

1 MetSerMetLeuValVal.....PheLeuLeuLeuTr 11
|||||:|||||:|||||:|||||:|||||:|||||:|
1282 ATGAGTTTCATGTCGACCTGAGTGCAAACTGCACTTCATGATGTTCTG 1331
11 pGlyVal1ThrTrpGly 16
|:|||||:|||||:|||||:|||||:|||||:|||||
1332 GAGGCTGACCTGGGGT 1347

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AA17774
seq_documentation_block:
ID AAX17774 standard; CDNA: 2394 BP.
XX
XX AAX17774:
AC
XX
XX 12-MAY-1999 (first entry)
DT
XX
XX YAK-1 related serine/threonine protein kinase-HTLAR33 encoding cDNA.
XX
XX YAK-1; serine-threonine protein kinase; HTLAR33; bone loss; ARDS;
KM inflammatory disorder; osteoporosis; Adult Respiratory Disease Syndrome;
KM arthritis; psoriasis; dermatitis; asthma; allergy; infection; HIV-1;
```

```
KW HIV-2; cachexia; immunodeficient disorder; septic shock; pain; injury;
KM cancer; anorexia; bulimia; Parkinson's disease; cardiovascular disease;
KM neurologic disorder; Huntington's disease; gene therapy; gene mapping;
KW Gilles de la Tourette's syndrome; ss.
XX
XX Homo sapiens.
OS
XX
XX EP894863-A1.
PN
XX
XX 03-FEB-1999.
PD
XX
XX 21-JUL-1998; 98EP-0305794.
PF
XX
XX 20-FEB-1998; 98US-0027064.
PR
XX 28-JUL-1997; 97US-0053924.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
PA
PI Bergsma DJ, Shabon U;
XX
XX WPI; 1999-108353/10.
DR
XX P-PSDB: AAW94997.
XX
XX New serine-threonine kinase (HTLAR33) polypeptides and
PT polynucleotides - useful as diagnostic reagents and for prevention
PT and treatment of bone loss, neurological and inflammatory disorders
XX
XX Claim 2; Page 13-15; 30pp; English.
XX
CC This cDNA encodes a YAK-1 related serine-threonine protein kinase
CC polypeptide (HTLAR33). Host cells transformed with an expression vector
CC comprising the HTLAR33 nucleic acid are used for the recombinant
CC production of the protein. HTLAR33 polynucleotides and polypeptides are
CC useful for are useful for diagnosing susceptibility to diseases and for
CC screening for antagonists, agonists. These can be used in treatment to
CC enhance (agonist) or block (antagonist) HTLAR33 activity. Diseases
CC diagnosed, prevented or treated include: bone loss and inflammatory
CC disorders including osteoporosis, Adult Respiratory Disease Syndrome
CC (ARDS), Rheumatoid arthritis, osteoarthritis, Inflammatory Bowel Disease
CC (IBD), psoriasis, dermatitis, asthma, allergies; bacterial, fungal,
CC protozoan and viral infections, especially those caused by HIV-1 or
CC HIV-2; HIV-associated cachexia and other immunodeficient disorders;
CC septic shock; pain; injury; cancers; anorexia; bulimia; Parkinson's
CC disease; cardiovascular disease including restenosis, atherosclerosis,
CC acute heart failure, myocardial infarction; hypotension; hypertension;
CC urinary retention; angina pectoris; ulcers; benign prostatic hypertrophy;
CC and psychotic and neurological disorders, including schizophrenia, manic
CC depression, anxiety; delirium, dementia, severe mental retardation and
CC dyskinesias, such as Huntington's disease or Gilles de la Tourette's
CC syndrome. HTLAR33 polypeptides are also useful for mapping genes to
CC chromosomes, allowing gene inheritance to be studied through linkage
XX analysis.
XX
SQ Sequence 2394 BP; 478 A; 801 C; 676 G; 439 T; 0 other;

alignment_scores:
    Quality: 46.00      Length: 14
    Ratio: 4.182        Gaps: 0
    Percent Similarity: 78.571    Percent Identity: 57.143

alignment_block:
US-09-471-276-831_COPY_1_16 x AAX17774 ..
Align seg 1/1 to: AAX17774 from: 1 to: 2394

2 SerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrp 15
|||||:|||||:|||||:|||||:|||||:|||||
1901 TCCCGCCTCTGATGACCTGCGACCTGCGGCGCTACCTGG 1942

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AA16647
```

```

seq_documentation_block:
ID AAA16647 standard; cDNA; 3191 BP.
XX
XX AAA16647:
XX
XX 16-JUN-2000 (first entry)
XX
XX Human secreted protein clone cK213_12 nucleotide sequence SEQ ID NO:59.
XX
XX Human; secreted protein; Immunestimulant; Immunesuppressant; viricide;
XX antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
XX antidiabetic; antisthmatic; antirheumatic; antiprotozoic;
XX antihypertoid; immune deficiency; severe combined immunodeficiency; SCID;
XX infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
XX connective tissue disease; multiple sclerosis; erythematosis;
XX rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
XX Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
XX insulin dependent diabetes mellitus; graft-versus-host disease;
XX autoimmune inflammatory eye disease; allergy; ss.
XX
XX Homo sapiens.
XX
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
XX 17-AUG-1998; 98US-0096815.
XX 04-SEP-1998; 98US-0099229.
XX 23-OCT-1998; 98US-0105368.
XX 08-JAN-1999; 99US-0115234.
XX 12-FEB-1999; 99US-0119931.
XX 18-FEB-1999; 99US-0120575.
XX 30-APR-1999; 99US-0132020.
XX 11-AUG-1999; 99US-0096622.
XX
XX (GENY ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
XX Moberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
XX Wong GG, Clark HF, Reichel K;
XX
XX WPI: 2000-205979/18.
XX P-PSDB: AAY94927.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
XX nutritional, chemokine, immune stimulating or suppressing,
XX hematopoiesis regulating, tissue growth, activin/inhibin
XX antiinflammatory or tumor inhibition activity -
XX
XX Claim 68: Page 530-531; 641pp: English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in
XX AA194698 to AA194960. Isolated from human adult brain, adult thyroid,
XX adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney
XX adult placenta, adult testis, whole embryo, adult cartilage, kidney,
XX foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
XX and adult bladder, cDNA libraries. The polynucleotides and proteins are
XX predicted to have biological activities which would make them suitable
XX for treating, preventing or ameliorating medical conditions in humans
XX and animals. The polynucleotides can be used as markers for tissues in
XX which the protein is preferentially expressed, as molecular weight
XX markers on Southern gels, and as chromosome markers or tags to identify
XX chromosomes or to map gene positions. The proteins can be used in the
XX treatment of immune deficiencies and disorders, such as severe combined
XX immunodeficiency (SCID), as well as viral, bacterial, fungal and other
XX infections. These infections include human immunodeficiency virus (HIV),
XX hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
XX candidiasis. The proteins can be used to treat autoimmune disorders such
XX as connective tissue disease, multiple sclerosis, systemic lupus
XX erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,

```

```

CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention.
XX
SQ Sequence 3191 BP: 998 A: 716 C: 835 G: 642 T: 0 other:
XX
alignment_scores:
XX      Quality: 46.00      Length: 13
XX      Ratio: 4.600      Gaps: 0
XX Percent Similarity: 76.923      Percent Identity: 61.538
XX
alignment_block:
US-09-471-276-831_COPY_1_16 x AAA16647 ..
XX
Align seg 1/1 to: AAA16647 from: 1 to: 3191
XX
4 LeuValValPheLeuLeuLeuTrpGlyValJThrTrpGly 16
XX      |||:::|||||::: ||| ||| |||||
XX      2840 TTAGCTGTTTTCATTTGTTGGTGTGTGGGCGGCGG 2878
XX
seq_name: /SID52/gcgcdata/geneseq/geneseq/NA2001.DAT:AAH54911
seq_documentation_block:
ID AAAH54911 standard: DNA: 3427 BP.
XX
AAH54911:
XX
03-SEP-2001 (first entry)
XX
5. epidermidis genomic polynucleotide sequence SEQ ID NO:4275.
XX
Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX
vaccination; endocarditis; ds.
XX
Staphylococcus epidermidis.
XX
WO200134809-A2.
XX
17-MAY-2001.
XX
09-NOV-2000; 2000WO-US30782.
XX
09-NOV-1999; 99US-0164258.
XX
(GLAX ) GLAXO GROUP LTD.
XX
Kimmerly WJ;
XX
WPI; 2001-316495/33.
XX
Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX
useful for vaccinating against infections, e.g. endocarditis -
XX
Claim 8; Page 2004-2005; 2188bp; English.
XX
AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX
(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX
(I) and (II) can have antibacterial activity and therefore can be used
XX
in vaccination. The nucleic acids (I) may be used to produce the
XX
5. epidermidis polypeptides (II) via the production of vectors
XX
containing them which are used to produce hosts cells which express the
XX
polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX
used to vaccinate subjects and to raise antibodies against the bacteria
XX
The polypeptides may also be used to assay for other inhibitors of the
XX
activity and therefore identify compounds that may be used for the
XX
treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX
AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX
polynucleotide sequences from the present invention. AAH55091 to
XX
AAH55098 represent oligonucleotide sequences and primers which are used
XX
in the exemplification of the present invention.

```


PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI: 2001-476164/51.
DR P-PSDB: AAM24394.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1: Page 726-728; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
XX
SQ Sequence 5345 BP; 1288 A; 1232 C; 1539 G; 1286 T; 0 other;

alignment_scores:
Quality: 46.00 Length: 22
Ratio: 3.067 Gaps: 1
Percent Similarity: 68.182 Percent Identity: 36.364

alignment_block:
US-09-471-276-831_COPY_1_16 x AAM99053 ..
Align seg 1/1 to: AAM99053 from: 1 to: 5345

1 MetSerMetLeuValVal.....PheLeuLeuTr 11
|||||:|||||:|||||:|||||:|||||:|||||
2397 ATGAGTTTGATGGTGACGTGAGCTGAACCTGCTCATGTTCTG 2446
11 pGlyValThrTrpGly 16
|:|||||:|||||:|||||:|||||:|||||
2447 GAGCGTGACCTGGGGT 2462

seq_name: /SIDS2/gcgcdata/geneseq/geneseqn/NA2000.DAT:AAZ34627

seq_documentation_block:
ID AAZ34627 standard; cDNA: 7478 BP.
XX
AC AAZ34627;
XX
DT 15-FEB-2000 (first entry)
XX
DE Mouse clock cDNA.
XX
KM Clock; mouse; transcription factor; circadian rhythm;
KW jet lag; sleep disorder; depression; seasonal affective disorder;
XX fertility; therapy; ss.
OS Mus musculus.
XX
XX
FH Key Location/Qualifiers
FT CDS 389..2956
FT /*tag= a
XX
XX
PN WO9957137-A1.
XX
PD 11-NOV-1999.
XX
PF 06-MAY-1999; 99WO-US10072.
XX
PR 07-MAY-1998; 98US-0084610.
XX
PA (HARD) HARVARD COLLEGE.
XX
PI Weltz CJ, Gekakis N, Staknis D;
XX

DR WPI: 2000-052938/04.
DR P-PSDB: AAY32214.
XX
PT Novel heterodimeric composition for identifying modulators used in
PT diagnosing and treating circadian clock disruption disorders -
XX
PS Disclosure: Fig 15A-B; 96pp; English.
XX
CC This is the nucleotide sequence of murine cDNA encoding CLOCK
CC protein (see AAY32214). CLOCK activates the transcription of the
CC perl gene when present in combination with BMAL1 (see AAY32209).
CC The invention, based on the discovery of the transcriptional
CC mechanism regulating genes responsible for the establishment and/or
CC maintenance of the circadian clock, is useful for the assay of
CC novel drugs aimed at restoration of a normal circadian cycle, the
CC drugs being modulators of BMAL1-CLOCK-mediated transcription of
CC E-box-linked genes. The drugs are used to treat conditions such as
CC jet lag, sleep disorders, depression (seasonal affective disorder)
CC and infertility. The invention also provides BMAL1 and CLOCK
CC proteins with which to stimulate the transcription of an
CC E-box-linked gene which regulates the circadian clock.
XX
SQ Sequence 7478 BP; 2177 A; 1534 C; 1616 G; 2151 T; 0 other;

alignment_scores:
Quality: 46.00 Length: 9
Ratio: 5.111 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-09-471-276-831_COPY_1_16 x AAZ34627 ..
Align seg 1/1 to: AAZ34627 from: 1 to: 7478

8 LeuLeuLeuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||:|||||:|||||
5598 CTCCTTCCTGATGCGGTGACCTTGGGT 5624

seq_name: /SIDS2/gcgcdata/geneseq/geneseqn/NA1998.DAT:AAV61401

seq_documentation_block:
ID AAV61401 standard; DNA: 7498 BP.
XX
AC AAV61401;
XX
DT 02-FEB-1999 (first entry)
XX
DE Mouse clock gene.
XX
KM Clock gene; circadian rhythm; mouse; jet lag; sleep-wake disorder;
KW seasonal affective disorder; cancer; diagnosis; therapy; ds.
XX
OS Mus sp.
XX
XX
FH Key Location/Qualifiers
FT CDS 389..2956
FT /*tag= a
XX
XX
PN WO9840514-A1.
XX
PD 17-SEP-1998.
XX
PF 13-MAR-1998; 98WO-US05114.
XX
PR 30-JUN-1997; 97US-0885291.
XX
PR 13-MAR-1997; 97US-0816693.
XX
PA (NOUN) UNIV NORTHWESTERN.
XX
XX
PI Pinto LH, Takahashi JS, Turek F;
XX
DR WPI: 1998-520828/44.

DR P-PSDB; AAW79529.
 XX Novel mouse and human circadian rhythm gene. Clock - useful for
 PT treating e.g. jet-lag, sleep-wake disorders, abnormal cell division,
 PT etc
 XX
 PS Claim 4; Fig 8.1-8.8; 154pp; English.
 XX
 CC This is the complete nucleotide sequence of the murine Clock gene
 CC based upon genomic exon sequences: splice donor and acceptor sites
 CC are provided in AAW61403-48. It encodes a 855-amino acid polypeptide
 CC (see AAW79529) that acts as a transcription factor. The Clock gene
 CC regulates at least 2 fundamental properties of the circadian clock
 CC system: the intrinsic circadian period and the persistence of
 CC circadian rhythmicity. Identification and isolation of the Clock
 CC gene involved: genetic mapping of the mouse genome (mid portion of
 CC chromosome 5); physical mapping of the Clock region; transcription
 CC unit analysis in the Clock region by (a) direct screening of SCN
 CC cDNA libraries with BAC clones as probes, (b) hybridisation
 CC selections of cDNAs from SCN libraries using BAC clones as driver,
 CC and (c) shutgun sequencing random M13 libraries made from BAC
 CC clones; transgenic mouse expression of an isolated BAC clone and
 CC phenotypic rescue of Clock⁻ and mRNA expression of candidate genes
 CC by Northern analysis in Clock mutant versus wild-type mice. The
 CC human Clock gene (see AAW61450) has also been identified. The
 CC invention provides isolated and purified CLOCK polypeptides. The
 CC polynucleotides (including antisense), vectors and host cells. The
 CC altered or disrupted circadian rhythms e.g. jet-lag, seasonal
 CC affective disorder, sleep-wake cycle disorders such as mood state,
 CC stress, neurological disorders, to regulate diet and food intake
 CC especially for diabetes, to treat cardiovascular, respiratory,
 CC liver or endocrine disorders, and for diagnosis and treatment of
 CC abnormal cell division such as cancer.
 XX
 SQ Sequence 7498 BP; 2197 A; 1534 C; 1616 G; 2151 T; 0 other:
 alignment_scores: Quality: 46.00 Length: 9
 Ratio: 5.111 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 88.889
 alignment_block:
 US-09-471-276-831_COPY_1_16 x AAW61401 ..
 Align seg 1/1 to: AAW61401 from: 1 to: 7498
 8 leuLeuLeuTrpGlyValThrTrpGly 16
 |||||:|||||:|||||:|||||:|||||
 5598 CTGCTTCTGTATGGCTGACTTGCGCT 5624
 seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2001.DAT:AAI21916
 seq_documentation_block:
 ID AAI21916 standard; DNA: 128 BP.
 AC AAI21916;
 XX
 DT 12-OCT-2001 (first entry)
 DE Probe #11849 for gene expression analysis in human cervical cell sample.
 KW Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 XX
 OS Homo sapiens.
 OS
 PN WO200157278-A2.
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00670.

XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-488901/53.
 XX
 PR Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 PS Claim 25; SEQ ID No 11849; 487pp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 128 BP; 18 A; 40 C; 37 G; 33 T; 0 other:
 alignment_scores: Quality: 45.00 Length: 11
 Ratio: 5.000 Gaps: 0
 Percent Similarity: 81.818 Percent Identity: 63.636
 alignment_block:
 US-09-471-276-831_COPY_1_16 x AAI21916 ..
 Align seg 1/1 to: AAI21916 from: 1 to: 128
 5 ValValPheLeuLeuLeuTrpGlyValThrTrp 15
 |||||:|||||:|||||:|||||:|||||
 74 GTCTCTTGTCTCTGCTGCGGCTGCGGTGG 106
 seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2001.DAT:AAI47201
 seq_documentation_block:
 ID AAI47201 standard; DNA: 128 BP.
 AC AAI47201;
 XX
 DT 17-OCT-2001 (first entry)
 DE Probe #15887 used to measure gene expression in human placenta sample.
 KW Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 OS
 PN WO200157272-A2.
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 15887; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 128 BP; 18 A; 40 C; 37 G; 33 T; 0 other;

alignment_scores:
Quality: 45.00 Length: 11
Ratio: 5.000 Gaps: 0
Percent Similarity: 81.818 Percent Identity: 63.636

alignment_block:

US-09-471-276-831_COPY_1_16 x AA147201 ..

Align seg 1/1 to: AA147201 from: 1 to: 128

5 ValValPheLeuLeuLeuTrpGlyValThrTrp 15
||||:||||:||||:||||:||||:||||
74 GTCCTCTTCTCCTGCTGCTGCGGTGTGCGGTGG 106

seq_name: /STDs2/gcgdata/geneseq/geneseqn/NA2001.DAT:AA107604

seq_documentation_block:

ID AA107604 standard; DNA; 128 BP.

XX
XX AC AA107604;

XX
XX DT 09-OCT-2001 (first entry)

XX
XX DE Probe #7595 used to measure gene expression in human breast sample.

XX
XX KW Probe; human; breast disease; breast cancer; development disorder; ss;
KM Inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX
XX OS Homo sapiens.

XX
XX PN WO200157270-A2.

XX
XX PD 09-AUG-2001.

XX
XX PF 29-JAN-2001; 2001WO-US00661.

XX
XX PR 04-FEB-2000; 2000US-0180312.

XX
XX PR 26-MAY-2000; 2000US-0207456.

XX
XX PR 30-JUN-2000; 2000US-0608408.

XX
XX PR 03-AUG-2000; 2000US-0632366.

XX
XX PR 21-SEP-2000; 2000US-0234687.

XX
XX PR 27-SEP-2000; 2000US-0236359.

XX
XX PR 04-OCT-2000; 2000GB-0024263.

XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-476286/51.
XX
XX PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
XX
XX Claim 25; SEQ ID No 7595; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast.
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 128 BP; 18 A; 40 C; 37 G; 33 T; 0 other;

alignment_scores:
Quality: 45.00 Length: 11
Ratio: 5.000 Gaps: 0
Percent Similarity: 81.818 Percent Identity: 63.636

alignment_block:

US-09-471-276-831_COPY_1_16 x AA107604 ..

Align seg 1/1 to: AA107604 from: 1 to: 128

5 ValValPheLeuLeuLeuTrpGlyValThrTrp 15
||||:||||:||||:||||:||||:||||
74 GTCCTCTTCTCCTGCTGCTGCGGTGTGCGGTGG 106

seq_name: /STDs2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA10193

seq_documentation_block:

ID AAA10193 standard; CDNA; 285 BP.

XX
XX AC AAA10193;

XX
XX DT 03-JUL-2000 (first entry)

XX
XX DE Rat liver toxicological response marker, SEQ ID NO:86.

XX
XX KW Toxicological response marker; rat; liver; expression pattern;

XX
XX KW Toxicity screening; toxic compound; polycyclic aromatic hydrocarbon;

XX
XX KW PAH; benzo(a)pyrene; clofibrate; acetaminophen; ss.

XX
XX OS Rattus norvegicus.

XX
XX PN WO200012760-A2.

XX
XX PD 09-MAR-2000.

XX
XX PF 27-AUG-1999; 99WO-US19768.

XX
XX PR 28-AUG-1998; 98US-0141825.

XX
XX PR 13-OCT-1998; 98US-0172108.

XX
XX PR 13-OCT-1998; 98US-0172711.

XX
XX PA (INCY-) INCYTE PHARM INC.

XX
XX PI Cunningham MJ, Zweiger GB, Panzer SR, Sellhammer JJ;

XX
XX DR WPI; 2000-237888/20.
XX
XX PT Isolated and purified nucleic acid molecules used as toxicological
PT response markers for detecting and diagnosing a potential toxicological

PT response in a mammalian subject to a test compound or molecule -
XX
XX
PS Claim 6: Page 65; 76pp: English.
XX
CC Sequences AAA10108-A10224 represent rat liver toxicological response
CC markers. These were identified by their pattern of at least twofold
CC upregulation or downregulation of expression in rat liver treated with
CC a toxic compound (e.g., clofibrate, acetaminophen or polycyclic
CC aromatic hydrocarbons (PAHs) such as benzo(a)pyrene). Fluorescently
CC labelled rat liver mRNA was contacted with a microarray comprising a
CC library of rat cDNA molecules. Twofold or larger changes in
CC hybridisation were only observed between the sample mRNA and sequences
CC AAA10108-A10224. In particular, sequences AAA10110, AAA10116, AAA10117,
CC AAA10120, AAA10126, AAA10133, AAA10138, AAA10140, AAA10142-A10144,
CC AAA10146, AAA10149, AAA10164, AAA10174, AAA10185, AAA10188, AAA10189,
CC AAA10201 and AAA10205 were all upregulated in samples treated with known
CC toxic compounds relative to untreated samples, while sequences AAA10150,
CC AAA10156, AAA10157, AAA10159-A10163, AAA10166, AAA10168, AAA10170,
CC AAA10175, AAA10178, AAA10181, AAA10192, AAA10194, AAA10197, AAA10202,
CC AAA10209, AAA10210, AAA10212 and AAA10222 were all downregulated.
CC Expression of these sequences is therefore modulated in liver during a
CC metabolic response to a toxic compound. The markers may be used as probes
CC to determine the toxicity of a test compound. A tissue sample from an
CC animal treated with the test compound is obtained, labelled (e.g., with a
CC fluorophore) and then contacted with a microarray comprising the markers.
CC The expression pattern of the markers may then be compared with the
CC marker expression pattern in untreated control samples, and the toxicity
CC of the test compound determined. The tissue sample is preferably selected
CC from liver, kidney, brain, spleen, pancreas and lung. The nucleic acid
CC molecules and methods of the invention may also be used for screening
CC libraries of molecules for specific binding affinity, and for the
CC fine-tuning of treatment regimens which use drugs with toxic side-effects
CC such that the side-effects are minimised without compromising the efficacy of the
XX
SQ Sequence 285 BP; 58 A; 78 C; 79 G; 70 T; 0 other;

alignment_scores: length: 16
Quality: 45.00 Gaps: 0
Ratio: 3.750
Percent Similarity: 75.000 Percent Identity: 56.250

alignment_block:
US-09-471-276-831_COPY_1_16 x AAA10193 ..
Align seg 1/1 to: AAA10193 from: 1 to: 285

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
|||||:||||:||||: |||||||||||||
14 ATGCTCTGCTGCTACTGACTGACTGCTGCTGCGGTTTCATTCTGGGC 61

seq_name: /SID52/gcgcdata/geneseq/geneseqn/NA2001.DAT:AA508124

seq_documentation_block:
ID AA508124 standard: cDNA; 285 BP.
XX
AC AA508124;
XX
DT 23-OCT-2001 (first entry)
XX
DE Mammalian toxicological response marker #13.
XX
KW Mammalian toxicological response marker; antigen; antibody; agonist; ss.
XX
OS Rattus norvegicus.
XX
PN WO200136684-A2.
XX
PD 25-MAY-2001.
XX
PF 16-NOV-2000; 2000WO-US31743.
XX
PR 19-NOV-1999; 99US-0443184.

XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Cunningham MJ, Zweiger GB, Kaser MR, Panzer SR, Sellhammer JJ;
PI Yue H, Baughn MR, Azimzal Y, Lal P;
XX
DR WPI: 2001-355646/37.
XX
PT Novel mammalian nucleic acid molecules whose levels are up regulated or
PT down regulated following treatment with a toxic compound, useful for
PT detecting metabolic and toxicological responses and in monitoring drug
PT action -
XX
PS Claim 1: Page 41; 89pp: English.
XX
CC The sequence is a novel mammalian nucleic acid molecule whose levels are
CC up regulated or down regulated following treatment with a toxic compound.
CC Polynucleotide sequences complementary to the sequences of the invention
CC are useful for preventing a toxicological response by acting against one
CC or more up-regulated nucleic acid molecules. An agonist identified by the
CC above method is also useful for preventing a toxicological response by
CC initiating transcription of a gene comprising a down regulated nucleic
CC acid molecule of the invention. Proteins encoded by the nucleic acids of
CC the invention are useful for producing an animal model system. The nucleic
CC acids immobilised on a substrate as hybridisable array element in a
CC microarray format may be used to characterise gene expression patterns
CC associated with novel compounds to elucidate any toxicological responses,
CC or to monitor the effects of treatments during clinical trials or therapy
CC where metabolic response to toxic compounds may be expected.
CC The nucleic acids are useful for various hybridisation technologies and
CC is useful for designing hybridisation probes. The nucleic acid molecule
CC or its fragment, or a protein encoded by the nucleic acid molecule may be
CC used to purify a ligand from a sample.
XX
SQ Sequence 285 BP; 58 A; 78 C; 79 G; 70 T; 0 other;

alignment_scores: length: 16
Quality: 45.00 Gaps: 0
Ratio: 3.750
Percent Similarity: 75.000 Percent Identity: 56.250

alignment_block:
US-09-471-276-831_COPY_1_16 x AA508124 ..
Align seg 1/1 to: AA508124 from: 1 to: 285

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
|||||:||||:||||: |||||||||||||
14 ATGCTCTGCTGCTACTGACTGACTGCTGCTGCGGTTTCATTCTGGGC 61

seq_name: /SID52/gcgcdata/geneseq/geneseqn/NA2001.DAT:AA112735

seq_documentation_block:
ID AA112735 standard: DNA; 472 BP.
XX
AC AA112735;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #2668 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX
KM cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.

PR 30-JUN-2000; 2000US-0608408

Penn SG, Hanzel DK, Chen W, Rank DR,

DR WPI: 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID No 2635; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 472 BP; 102 A; 121 C; 145 G; 104 T; 0 other;

alignment_scores:
Quality: 45.00 Length: 11
Ratio: 5.000 Gaps: 0
Percent Similarity: 81.818 Percent Identity: 63.636

alignment_block:
US-09-471-276-831_COPY_1_16 x AA102644 ..
Align seg 1/1 to: AA102644 from: 1 to: 472

5 ValValPheLeuLeuLeuTrpGlyValThrTrp 15
|||||:|||||:|||||:|||||
380 GTCCCTCTTCTCTCTGCTGCGGTGCGTGG 412

seq_name: /SID52/gcdata/geneseq/geneseqn/NA2001.DAT:AAH10448

seq_documentation_block:
ID AAH10448 standard; CDNA: 576 BP.
AC AAH10448;
XX
XX 26-JUN-2001 (first entry)
DE Human cDNA clone (3'-primer) SEQ ID NO:7283.
XX
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
OS
PM EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3; SEQ ID 7283; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 576 BP; 116 A; 153 C; 148 G; 156 T; 3 other;

alignment_scores:
Quality: 45.00 Length: 14
Ratio: 3.750 Gaps: 0
Percent Similarity: 85.714 Percent Identity: 42.857

alignment_block:
US-09-471-276-831_COPY_1_16 x AAH10448 ..
Align seg 1/1 to: AAH10448 from: 1 to: 576

2 SerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrp 15
:|||||:|||||:|||||:|||||
283 AACCTCTTAACCTCTTTTGTGATTTTGTGGGAAACATCTCG 324

seq_name: /SID52/gcdata/geneseq/geneseqn/NA2000.DAT:AAF10038

seq_documentation_block:
ID AAF10038 standard; CDNA: 632 BP.
AC AAF10038;
XX
XX 13-MAR-2001 (first entry)
DE Fusarium venenatum EST SEQ ID NO:2561.
XX
XX
XX Fusarium venenatum
OS
XX
XX WO200056762-A2.
PM
XX
XX 28-SEP-2000.
PD
XX
XX 22-MAR-2000; 2000WO-US07781.
PR
XX
XX 22-MAR-1999; 99US-0273623.
PR
XX

Multiple gene expression; filamentous fungal cell; EST;
expressed sequence tag; Fusarium venenatum; Aspergillus niger;
Aspergillus oryzae; Trichoderma reesei; identification; recombination;
culture condition; environmental stress; spore morphogenesis;
metabolic pathway engineering; catabolic pathway engineering; ss.

PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI: 2000-594572/56.
DR
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 86: Page 1320; 3161pp: English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from *Aspergillus*
CC *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
CC all specifically claimed in the present invention.
XX
S0 Sequence 632 BP; 178 A; 175 C; 118 G; 156 T; 5 other;
XX
alignment_scores:
Quality: 45.00 Length: 15
Ratio: 4.500 Gaps: 0
Percent Similarity: 66.667 Percent Identity: 46.667
alignment_block:
US-09-471-276-831_COPY_1_16 x AAF10038/rev ..
Align seg 1/1 to reverse of: AAF10038 from: 1 to: 632
2 SetMetLeuValAlaPheLeuLeuLeuTrpValaThrTrpGly 16
||| :||||| ||||||||| :|||
233 TCAGCGCTGATGATTTCAAAAGACGTATGGGGAANAAGCTGGGCG 189
seq_name: /SIDSz/gcgcdata/geneseq/geneseqn/NA1999.DAT:AAx24234
seq_documentation_block:
ID AAX24234 standard: DNA; 748 BP.
XX
XX AAX24234;
XX
DT 01-JUL-1999 (first entry)
XX
DE WO9916900 Seq ID 4.
XX
XX Gram positive bacteria; covalent coupling; pathogenicity factors
XX LPxNG-motif dependent C-terminal anchorage; cell surface; antibacterial;
XX therapeutic agent; extracellular matrix; ss.
OS Unidentified.
XX
XX WO9916900-A1.
PN

```

XX 08-APR-1999.
PD
XX
XX 26-SEP-1998; 98MO-EP06136.
PF
XX 29-OCT-1997; 97EP-0118755.
PR
XX 27-SEP-1997; 97EP-0116841.
PR
XX (EVOT-) EVOTEC BIOSYSTEMS AG.
PA
XX
XX Goetz F, Pohlner J, Strauss A, Thumm G;
PI
XX WPI: 1999-255104/21.
DR
XX
XX Identifying nucleic acid that affects adhesion of Gram positive
PT bacteria
PR
XX
XX Claim 17; Page 41-42; 58pp; German.
PS
XX
XX This invention describes a novel method for identifying a nucleic acid
CC that encodes a polypeptide which affects covalent coupling of
CC polypeptides to the surface of Gram-positive bacteria. The method
CC identifies mutations in bacterial factors that (in)directly interact
CC with the LPXTG-motif dependent C-terminal anchorage of polypeptides,
CC particularly pathogenicity factors, to the cell surface. The products
CC of the invention have antibacterial activity and are potentially
CC useful as therapeutic agents, inhibiting binding of bacteria to
CC extracellular matrix.
CC This sequence is represented in the Seq ID listing but does not
CC appear in the specification.
CC
XX
XX Sequence 748 BP; 216 A; 113 C; 170 G; 238 T; 11 other;
SO
XX
XX
XX alignment_scores:
XX      Quality: 45.00      Length: 8
XX      Ratio: 6.429      Gaps: 0
XX      Percent Similarity: 87.500      Percent Identity: 87.500
XX
XX alignment block:
XX US-09-471-276-831_COPY_1_16 x AAX24234 ..
XX
XX Align seg 1/1 to: AAX24234 from: 1 to: 748
XX
XX      9 LeuLeuTrpGlyValThrTrpGly 16
XX      |||||
XX      585 CTTTGTGTGGGGCCCACTGGGGGT 608
XX
XX seq_name: /SID52/gcdata/geneseq/geneseq/NA1999.DAT:AAX24250
XX
XX seq_documentation block:
XX ID AAX24250 standard; DNA: 748 BP.
XX
XX AAX24250;
XX AC
XX
XX 01-JUL-1999 (first entry)
DT
XX
XX WO9916894 Seq ID 4.
DE
XX
XX Active agent; covalent coupling; Gram-positive bacteria; inhibitor;
XX adhesion; extracellular matrix; LPXTG motif-dependent anchorage;
XX pathogenicity factor; bacterial cell wall; therapeutic agent; ss.
XX
XX Unidentified.
OS
XX
XX WO9916894-A1.
PN
XX
XX 08-APR-1999.
PD
XX
XX 26-SEP-1998; 98MO-EP06137.
PF
XX
XX 29-OCT-1997; 97EP-0118756.
PR
XX 27-SEP-1997; 97EP-0116840.
PR

```

XX (EVOT-) EVOTEC BIOSYSTEMS AG.
PA Goetz F, Pohlner J, Strauss A, Thumm G;
PI WPI: 1999-255102/21.
XX Identifying agents that affect adhesion of Gram positive bacteria
XX
XX Disclosure: Page 38-39; 51pp; German.
XX This invention describe a novel method for identifying an active agent
CC that affects covalent coupling of polypeptides to the surface of
CC Gram-positive bacteria. The agents identified inhibit
CC polypeptide-induced adhesion of bacteria to extracellular matrix in the
CC host. The active agent effects the LPXTG motif-dependent anchorage of
CC pathogenitly factors to the bacterial cell wall and are potentially
CC useful as therapeutic agents, inhibiting binding of bacteria to
CC extracellular matrix in the host.
CC This sequence is not described in the specification but is represented
CC in the sequence ID listing.
XX
SQ Sequence 748 BP; 216 A; 113 C; 170 G; 238 T; 11 other;

alignment_scores:
Quality: 45.00 Length: 8
Ratio: 6.429 Gaps: 0
Percent Similarity: 87.500 Percent Identity: 87.500

alignment_block:

US-09-471-276-831_COPY_1_16 x AAX24250 ..

Align seg 1/1 to: AAX24250 from: 1 to: 748

9 LeuLeuTrpGlyValThrTrpGly 16

|||||
585 CTTTGTGGGCCCCACCTGGCGT 608

seq_name: /SID52/gcdata/geneseq/geneseq/NA1995.DAT:AA087927

seq_documentation_block:

ID AA087927 standard: DNA; 794 BP.

XX AA087927:

XX 22-NOV-1995 (first entry)

XX Anther specific gene promoter element.

XX Anther specific gene: Brassica napus; antisense RNA; Cruciferae;

XX pollen formation; male sterile plant; ss.

XX Brassica napus.

XX JP07059573-A.

XX 07-MAR-1995.

XX 20-AUG-1993; 93JP-0206459.

XX 20-AUG-1993; 93JP-0206459.

XX (MITS) MITSUBISHI CORP.

XX (MITU) MITSUBISHI KASEI CORP.

XX WPI: 1995-135897/18.

XX A gene expressed specifically in anther(s) - used for the
XX preparation of male sterile Cruciferae plants

XX Claim 5; Page 5-6; 6pp; Japanese.

CC This sequence was used for the expression of an anther specific gene
CC derived from Brassica napus. This sequence caused expression of
CC the anther specific gene as antisense RNA in Cruciferae plant causing
CC pollen formation to be modified. This is esp. useful in the production
CC of male sterile plants.
XX

SQ Sequence 794 BP; 258 A; 157 C; 137 G; 228 T; 14 other;

alignment_scores:
Quality: 45.00 Length: 15
Ratio: 3.750 Gaps: 0
Percent Similarity: 80.000 Percent Identity: 46.667

alignment_block:

US-09-471-276-831_COPY_1_16 x AA087927/rev ..

Align seg 1/1 to reverse of: AA087927 from: 1 to: 794

2 SerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrpGly 16

||| :|||:||||| :|||:|||||
763 TCTTGCTTGTGCATATTTTGTGTGTTATGGCCTTTTATGGGG 719

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2002, 16:45:58 ; Search time 16.52 Seconds

(without alignments)
279.647 Million cell updates/sec

Title: US-09-471-276-831

Perfect score: 661
Sequence: 1 MSMLVFFLLMGVTWGPVTE.....LLELTGPKVACSLADGAS 126

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SWISSProt_39:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	325	49.2	474	1	ALBG_HUMAN
2	74	11.2	192	2	VIF_HV1RH
3	73	11.0	663	1	MMO2_CHICK
4	71	10.7	404	1	FCG1_MOUSE
5	69	10.4	192	1	VIF_HV1UR
6	68.5	10.4	244	1	OSL3_ARATH
7	67	10.1	362	1	IBS5_HUMAN
8	67	10.1	609	1	Y4PA_RHISN
9	66.5	10.1	443	1	HEM1_HELMO
10	66.5	10.1	503	1	IRF7_HUMAN
11	66.5	10.1	5065	1	EPPL_HUMAN
12	66	10.0	362	1	IB16_HUMAN
13	66	10.0	2482	1	VWF_PIG
14	65	9.8	307	1	BLAC_MYCTU
15	65	9.8	483	1	VP19_VZVD
16	65	9.8	504	1	AFY1_HUMAN
17	64.5	9.8	258	1	RSEA_BACSU
18	64	9.7	192	1	VIF_HV1MA
19	64	9.7	1069	1	S24B_ARATH
20	64	9.6	1085	1	BUB1_HUMAN
21	63.5	9.5	1669	1	CA14_HUMAN
22	63	9.5	366	1	IC12_HUMAN
23	63	9.5	366	1	IC13_HUMAN
24	63	9.5	366	1	IC14_HUMAN
25	63	9.5	470	1	NRAM_IQOIT
26	63	9.5	752	1	PARC_ECOLI
27	63	9.5	1090	1	GUXB_CELFI
28	62.5	9.5	508	1	CP2B_HUMAN
29	62.5	9.5	1669	1	CA14_MOUSE
30	62	9.4	362	1	IB25_HUMAN
31	62	9.4	366	1	IC04_HUMAN
32	62	9.4	366	1	IC15_HUMAN
33	62	9.4	620	1	GG95_HUMAN

34	62	9.4	971	1	Y029_HUMAN	Q15032 homo sapien
35	62	9.4	1020	1	ACAL_ARATH	Q37145 arabidopsis
36	62	9.4	1247	1	AL71_HUMAN	Q91V69 arabidopsis
37	61.5	9.3	291	1	DM43_DIDNA	P82957 didephlis m
38	61.5	9.3	591	1	EVY1_MOUSE	P97767 mus musculu
39	61.5	9.3	1142	1	ENAM_PIG	O97939 sus scrofa
40	61.5	9.3	3591	1	FHAB_BORPE	P12255 bordelella
41	61	9.2	192	1	VIF_HV1A2	P03402 human immun
42	61	9.2	192	1	VIF_HV1N5	P12504 human immun
43	61	9.2	309	1	GAG_FUJVS	P03326 fujiinami sa
44	61	9.2	323	1	PIP_STRCO	Q98214 streptomyce
45	61	9.2	348	1	FTSY_MYCPN	P75362 mycoplasma
46	61	9.2	362	1	IB49_HUMAN	P18464 homo sapien
47	61	9.2	544	1	IC43_BOVIN	Q28125 bos taurus
48	61	9.2	599	1	ESR1_MOUSE	P19785 mus musculu
49	61	9.2	685	1	SNK_HUMAN	Q9nyh3 homo sapien
50	61	9.2	1572	1	BA12_HUMAN	O60241 homo sapien
51	61	9.2	2051	1	FAS1_YEAS	P07149 s fatty aci
52	61	9.2	3567	1	ERY2_SACER	Q03132 saccharopol
53	60.5	9.2	477	1	PROX_MOUSE	P51175 mus musculu
54	60.5	9.2	976	1	EPAL_HUMAN	P21709 homo sapien
55	60.5	9.2	1107	1	YJEP_ECOLI	P39285 escherichia
56	60	9.1	192	1	VIF_HV1OY	P20890 human immun
57	60	9.1	360	1	CCD2_CAEEL	P35799 caenorhabdi
58	60	9.1	361	1	IB14_HUMAN	P03989 homo sapien
59	60	9.1	362	1	IB18_HUMAN	P10318 homo sapien
60	60	9.1	362	1	IB19_HUMAN	O08136 homo sapien
61	60	9.1	362	1	IB63_HUMAN	P08136 homo sapien
62	60	9.1	366	1	IC11_HUMAN	Q29631 homo sapien
63	60	9.1	369	1	MALK_SALTY	P19566 salmonella
64	60	9.1	547	1	YAGM_ECOLI	P77694 escherichia
65	60	9.1	611	1	ACES_FELCA	O62763 felis silve
66	60	9.1	752	1	PARC_SALTY	P26973 salmonella
67	60	9.1	912	1	TRX2_HUMAN	Q28062 bos taurus
68	60	9.1	2715	1	GAG_AVEV1	Q9umh6 homo sapien
69	59.5	9.0	239	1	MTGA_KLEPN	P06936 avian endog
70	59.5	9.0	242	1	PIV2_ADEO2	O48465 klebsiella
71	59.5	9.0	449	1	EVY1_HUMAN	P03372 human adeno
72	59.5	9.0	592	1	AHMK_HUMAN	O99502 homo sapien
73	59.5	9.0	2960	1	DYHC_EMENT	O09666 homo sapien
74	59.5	9.0	4344	1	INL3_RAT	P45444 emericella
75	59	8.9	105	1	ANP_HEHAM	O9wuk0 rattus norv
76	59	8.9	163	1	VIF_HV1Z6	P05140 hemitriptier
77	59	8.9	188	1	VIF_HV1ND	P04596 human immun
78	59	8.9	192	1	RL25_MYCLE	P18805 human immun
79	59	8.9	215	1	DT3E_PSECI	O9cd48 mycobacteri
80	59	8.9	290	1	FTSY_MYCGE	O50840 pseudomonas
81	59	8.9	346	1	IB40_HUMAN	P10320 mycoplasma
82	59	8.9	359	1	IB41_HUMAN	P10320 homo sapien
83	59	8.9	362	1	IB42_HUMAN	P30482 homo sapien
84	59	8.9	362	1	IB45_HUMAN	P30482 homo sapien
85	59	8.9	362	1	IB45_HUMAN	P30485 homo sapien
86	59	8.9	362	1	IB54_HUMAN	P30491 homo sapien
87	59	8.9	362	1	IB59_HUMAN	P30491 homo sapien
88	59	8.9	366	1	IC03_GORGO	P30386 gorilla gor
89	59	8.9	366	1	IC04_GORGO	P30386 gorilla gor
90	59	8.9	366	1	IC07_HUMAN	P30304 homo sapien
91	59	8.9	501	1	CP2B_RAT	O35132 rattus norv
92	59	8.9	537	1	DEXB_STRCO	O59905 streptococc
93	59	8.9	553	1	MIS_RAT	P49000 rattus norv
94	59	8.9	1538	1	PPSB_MYCTU	O10978 mycobacteri
95	59	8.9	4563	1	APB_HUMAN	P04114 homo sapien
96	58.5	8.9	101	1	FHE3_RAT	O63550 rattus norv
97	58.5	8.9	288	1	FHE3_MOUSE	O9qy14 mus musculu
98	58.5	8.9	319	1	FXE3_HUMAN	O13661 homo sapien
99	58.5	8.9	601	1	LONH_HAEIN	P43665 haemophilus
100	58.5	8.9	662	1	PD11_MOUSE	Q92185 mus musculu

RESULT 1

ALIGNMENTS

```

A1BG_HUMAN          STANDARD:          PRT:          474 AA.
ID   A1BG_HUMAN
AC   P04217
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ALPHA-1B-GLYCOPROTEIN.
GN   A1BG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX   NCBI_TaxId=9606;
RN   1
RP   SEQUENCE. DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.
RX   MEDLINE:86505955; PubMed:3458201;
RA   Ishiko N., Takahashi N., Putnam F.W.;
RT   *Amino acid sequence of human plasma alpha 1b-glycoprotein: homology
RT   to the immunoglobulin supergene family.*
RL   Proc. Natl. Acad. Sci. U.S.A. 83:2363-2367(1986).
CC   -1- FUNCTION: PLASMA GLYCOPROTEIN OF UNKNOWN FUNCTION.
CC   -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
DR   PIR: A02113; OM01B.
DR   SWISS-2DPAGE: P04217; HUMAN.
DR   MIM: 138670; -.
DR   InterPro: IPR003006; Ig_MHC.
DR   InterPro: IPR003598; Ig_C2.
DR   InterPro: IPR003600; Ig_Like.
DR   Pfam: PF00047; Ig_4.
DR   SMART: SM00408; IGC2; 1.
DR   SMART: SM00410; IG_Like; 4.
KM   Immunoglobulin domain; Glycoprotein; Plasma.
FT   DOMAIN          1   92   IG-LIKE V-TYPE DOMAIN 1.
FT   DOMAIN          93  185   IG-LIKE V-TYPE DOMAIN 2.
FT   DOMAIN         186  278   IG-LIKE V-TYPE DOMAIN 3.
FT   DOMAIN         279  376   IG-LIKE V-TYPE DOMAIN 4.
FT   DOMAIN         377  474   IG-LIKE V-TYPE DOMAIN 5.
FT   DISULFID        28   72   PROBABLE.
FT   DISULFID        402  449   PROBABLE.
FT   DISULFID        118  161   PROBABLE.
FT   DISULFID        211  258   .
FT   DISULFID        304  353   .
FT   CARBOHYD        23   23   N-LINKED (GLCNAC. . .).
FT   CARBOHYD        158  158   N-LINKED (GLCNAC. . .).
FT   CARBOHYD        342  342   N-LINKED (GLCNAC. . .).
FT   CARBOHYD        350  350   N-LINKED (GLCNAC. . .).
SQ   SEQUENCE          474 AA: 51940 MW: 48625472A920FA CRC64;

```

Query Match 49.2%; Score 325; DB 1; Length 474;

Best Local Similarity 71.7%; Pred. No. 4e-27; Indels 8; Gaps 3;

Matches 71; Conservative 3; Mismatches 17; Indels 8; Gaps 3;

```

QY 22 AIFETQXSLMAESEXHLTKLCO---CDADVPGPGDSRLPAVDQWGAQEPVHLDSPAI 77
DB 1 AIFETQPSLMAESEXHLTKLCOAVNTLTQARLETP--DFQL--FKNGVAQEPVHLDSPAI 56
QY 78 KHQFLTGTQGRYRCRSLSTGCMXOLSKLLETGPKVL 116
DB 57 KHQFLTGTQGRYRCRSLSTGCMXOLSKLLETGPKSL 95

```

RESULT 2

VIF_HVIRH STANDARD: PRT: 192 AA.

```

AC P05900;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION INFECTIVITY FACTOR (SOR PROTEIN).
GN VIF.
OS Human immunodeficiency virus type 1 (RF/HT isolate) (HIV-1).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

```

```

OX NCBI_TaxId=11701;
RN 1
RP SEQUENCE FROM N.A.
RA Starcich B.R., Hahn B.H., Shaw G.M., McNeely P.D., Modrow S.,
RA Wolf H., Parks E.S., Parks W.P., Josephs S.F., Gallo R.C.,
RA Wong-Staal F.;
RL Submitted (XXI-1987) to the HIV data bank.
CC -1- FUNCTION: DETERMINES VIRUS INFECTIVITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M17451; AAA45054.1; -.
DR HIV: M17451; VIFSRF.
DR InterPro: IPR000475; Viral_infect.
DR Pfam: PF00559; VIF; 1.
DR PRINTS: PR00349; VIRIONINFECT.
DR ProDom: PD000063; Viral_infect; 1.
KW AIDS.
SQ SEQUENCE          192 AA: 22687 MW: F005E0AE621A5C6C CRC64;

```

Query Match 11.2%; Score 74; DB 1; Length 192;

Best Local Similarity 22.0%; Pred. No. 0.69;

Matches 22; Conservative 14; Mismatches 40; Indels 24; Gaps 3;

```

QY 8 LLLMGV-----TWGVTFAAIFETQXSLMAESEXHLTKLCCDADVPGPGDSRLPAV 61
DB 8 MIVQVDRMRIRTKSLVKNHMYISRKAKWGYRNHYESTNRISSEVNIIPRODERLYIT 67
QY 62 QENGAQEPVHLDSPAIKHQFLTGTQGRYRCRSLSTGCM 101
DB 68 TYWG-----LHTGERD--WHLGGGVSIEM 89

```

RESULT 3

MM02_CHICK STANDARD: PRT: 663 AA.

```

AC 090611;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 72 KDA TYPE IV COLLAGENASE PRECURSOR (EC 3.4.24.24) (72 KDA
DE GELATINASE) (MATRIX METALLOPROTEINASE-2) (MMP-2) (GELATINASE A).
GN MMP2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxId=9031;
RN 1
RP SEQUENCE FROM N.A.
RC TISSUE: Embryo.
RX MEDLINE:94280397; PubMed:8010954;
RX Aimes R.T., French D.L., Quigley J.P.;
RT *Cloning of a 72 kDa matrix metalloproteinase (gelatinase) from
RT chicken embryo fibroblasts using gene family PCR: expression of the
RT gelatinase increases upon malignant transformation.*;
RL Biochem. J. 300:729-736(1994).

```

SEQUENCE OF 27-41 AND 107-122.

MEDLINE:91161603; PubMed:1848240;

Chen J.-M., Aimes R.T., Ward G.R., Youngielb G.L., Quigley J.P.;

*Isolation and characterization of a 70-kDa metalloproteinase

(gelatinase) that is elevated in Rous sarcoma virus-transformed


```

RP SEQUENCE FROM N.A.
RX MEDLINE-92148136: PubMed-1737933:
RA Hildebrand W.H., Madrigal J.A., Little A.-M., Parham P.;
RT "HLA-B*22: a family of molecules with identity to HLA-B7 in the alpha
  1-helix."
RL J. Immunol. 148:1155-1162(1992).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77774: AAA03686.1: -.
DR HSSP: P30685: 1A1N.
DR MIM: 142830: -.
DR InterPro: IPR003006: Ig_MHC.
DR InterPro: IPR003597: Ig_C1.
DR InterPro: IPR01039: MHC_I.
DR Pfam: PF00047: Ig: 1.
DR Pfam: PF00129: MHC_I: 1.
DR ProDom: PD000050: MHC_I: 1.
DR SMART: SM00407: IG_C1: 1.
DR PROSITE: PS00290: IG_MHC: 1.
DR MHC I: Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 362
FT DOMAIN 25 114 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 115 206 EXTRACTED ALPHA-1.
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.
FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.
FT TRANSMEM 309 332 CONNECTING PEPTIDE.
FT DOMAIN 333 362 CYTOPLASMIC TAIL.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 125 188 BY SIMILARITY.
FT DISULFID 227 283 BY SIMILARITY.
SQ SEQUENCE 362 AA: 40380 MW: 111730024578515D CRC64:

Query Match 10.1%; Score 67; DB 1; Length 362;
Best Local Similarity 24.2%; Pred. No. 7.9;
Matches 32; Conservative 14; Mismatches 64; Indels 22; Gaps 5;

QY 6 VFLLLMGV-----TWGPVTEALIFETQXSLMASEHXLKTLG-----OCDADVCP 52
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
DB 9 LLLLLLGGALALLETWAGSHSMRYFY-TAMSRPGRGEPRIAVGVYDDTQFVRPDSASP 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 53 PGDSRLPAVQEMGAQEPVHLDSPAIKHQFLITDGTGRYRCRGLSTGMXOLSKLELTG 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 68 RGPAPRPWVQEG---PEYWDNRTQITKAOQTDRSLRNLK-----GYYNQSEAGSHTW 119
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 113 PKVLASLADG 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 120 QIMYGCDLGPDG 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
Y4PA.RHISN STANDARD: PRT; 609 AA.
AC P55610:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE PUTATIVE TRANSCRIPTIONAL REGULATORY PROTEIN Y4PA.
GN Y4PA.
OS Rhizobium sp. (strain NGR234).

```

```

OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97305956: PubMed-9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
  Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes."
RL Nature 387:394-401(1997).
CC -1- FUNCTION: PROBABLE TRANSCRIPTIONAL REGULATOR THAT ACTS IN
CC CONJUNCTION WITH SIGMA-54.
CC -1- SIMILARITY: THE CENTRAL REGION CONTAINS A SIGMA-54 FACTOR
CC INTERACTION ATP-BINDING DOMAIN, TO Y4OT AND Y4OV.
CC -1- SIMILARITY: IN THE N-TERMINAL, TO Y4OT AND Y4OV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000089: AAB91811.1: -.
DR InterPro: IPR002197: HTH_Fts.
DR InterPro: IPR002078: Sig54_interact.
DR Pfam: PF00158: sigma54; 2.
DR PROSITE: PS00675: SIGMA54_INTERACT_1; FALSE NEG.
DR PROSITE: PS00676: SIGMA54_INTERACT_2; FALSE NEG.
DR PROSITE: PS00688: SIGMA54_INTERACT_3; FALSE NEG.
DR PROSITE: PS50045: SIGMA54_INTERACT_4; 1.
DR PROSITE: PS50045: SIGMA54_INTERACT_4; 1.
DR Hypothetical protein; Transcription regulation; DNA-binding;
  ATP-binding; Plasmid.
FT DOMAIN 313 533 SIGMA-54 FACTOR INTERACTION (POTENTIAL).
FT NP_BIND 395 404 ATP (POTENTIAL).
FT DNA_BIND 578 597 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 609 AA: 66700 MW: 8CC727E67D508F36 CRC64:

Query Match 10.1%; Score 67; DB 1; Length 609;
Best Local Similarity 30.3%; Pred. No. 14;
Matches 23; Conservative 9; Mismatches 18; Indels 26; Gaps 5;

QY 32 WASEHXLKTLGGCDAVDVGP-PGDSRLPAVQEMGAQEPVHLDSPAIKHQFLITDGTGR 90
   : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
DB 539 WIKSEH-LPPLGDANADAPHPHGEER-----EW-----IL--DALGR 573
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 91 YKCRSLSTGMXOLSK 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 574 HRFRRGEAARYLGISR 589
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
HEM1_HELMO STANDARD: PRT; 443 AA.
ID HEM1_HELMO
AC Q9ZG66;
DT 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE GLUTAMYL-TRNA REDUCTASE (EC 1.2.1.-) (GLUTR).
GN HEMA.
OS Helobacillus mobilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Helobacterium group; Helobacillus.
OX NCBI_TaxID=28064;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99061957: PubMed-9843979;
RA Xiong J., Inoue K., Bauer C.E.;
RT "Tracking molecular evolution of photosynthesis by characterization of
  a major photosynthesis gene cluster from Helobacillus mobilis."

```

RL Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
 CC -1- CATALYTIC ACTIVITY: GLUTAMYL-TRNA(GLU) + NADPH --> GLUTAMATE-1-
 CC SEMIALDEHYDE + NADP(+) + TRNA(GLU).
 CC -1- PATHWAY: FIRST STEP IN PORPHYRIN BIOSYNTHESIS BY THE C5 PATHWAY.
 CC -1- SIMILARITY: BELONGS TO THE GLUTAMYL-TRNA REDUCTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF080002; AAC84013.1; -.
 DR HSSP: Q42843; 1B29.
 DR InterPro: IPR000343; Glutr.
 DR Pfam: PF00745; Glutr. 1.
 DR PROSITE: PS00747; Glutr. 1.
 DR Porphyryn biosynthesis; Oxidoreductase; NADP-
 KM Porphyryn biosynthesis; Oxidoreductase; NADP-
 SO SEQUENCE 443 AA; 49577 MW; E95C30E251A0C5F6 CRC64;

Query Match 10.1%; Score 66.5; DB 1; Length 443;
 Best Local Similarity 26.2%; Pred. No. 11;
 Matches 34; Conservative 15; Mismatches 34; Indels 47; Gaps 9;

OY 18 VTEALIFVETQXSLMAESENHLK-----TLGGCD-ADVCPGPGDSR--LPVQ--- 62
 DB 17 VREKLSFTFAQLS---EALHKLDGMAGIBGCCILSTGCTNRTETYGASTDEKGMTAVKREV 73
 OY 63 -EWGAEPR-----VHLDSPAIKHOF-----LITGDTQ--GRYR-----C 93
 DB 74 LEWGLQGPQDFSKYFVHTLYDAIRHLFRVASGLDSMWLGEGTILQVRYAVRGSCNEQC 133
 OY 94 RGLSLTGMXQ 103
 DB 134 SNGIVNTMFO 143

RESULT 10
 IRF7_HUMAN STANDARD: PRT; 503 AA.
 ID IRF7_HUMAN
 AC Q92985; O00331; O00332; O00333; O75924;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE INTERFERON REGULATORY FACTOR 7 (IRF-7).
 GN IRF7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Plimates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RC TISSUE: Spleen;
 RA Grossman A., Nicholl J., Antonio L., Luethy R., Suggs S.,
 RA Sutherland G.R., Mak T.W.;
 RL Submitted (Oct-1996) to the EMBL/Genbank/DBD databases.
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
 RA Zhang L., Pagano J.S.;
 RA MEDLINE:97459673; PubMed:9315633;
 RT IRF-7, a new interferon regulatory factor associated with Epstein-
 RT Barr virus latency.*;
 RL Mol. Cell. Biol. 17:5748-5757(1997).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM D).
 RX MEDLINE:99003279; PubMed:9786932;
 RA Au W.-C., Moore P.A., Lafleur D.W., Tombal B., Pitba P.M.;
 RT Characterization of the interferon regulatory factor-7 and its
 RT potential role in the transcription activation of interferon A
 genes.*;
 RT

RL J. Biol. Chem. 273:29210-29217(1998).
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. BINDS TO THE INTERFERON-
 CC STIMULATED RESPONSE ELEMENT (ISRE) IN IRF PROMOTERS AND IN THE O
 CC PROMOTER (OP) OF EBV NUCLEAR ANTIGEN-1 (EBNA1).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), B/BETA, C/GAMMA
 CC AND D/H; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN SPLEEN, THYMUS, AND
 CC PERIPHERAL BLOOD LEUCOCYTES.
 CC -1- SIMILARITY: BELONGS TO THE IRF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: 073036; AAB17190.1; -.
 DR EMBL: 053830; AAB80686.1; -.
 DR EMBL: 053831; AAB80688.1; -.
 DR EMBL: 053832; AAB80690.1; -.
 DR EMBL: AF076494; AAC70999.1; -.
 DR HSSP: P23906; 1IRC.
 DR MIM: 605047; -.
 DR InterPro: IPR001346; IRF.
 DR Pfam: PF00605; IRF. 1.
 DR PRINTS: PR00267; INTERNREGCT.
 DR PRODOM: PD002355; IRF. 1.
 DR SMART: SM00348; IRF. 1.
 DR PROSITE: PS00601; IRF. 1.
 KW Transcription regulation; DNA-binding; Nuclear protein; Activator;
 KM Alternative splicing.
 FT DNA-BIND 13 122
 FT VARSPLIC 1 6
 FT VARSPLIC 152 164
 FT VARSPLIC 165 256
 FT VARSPLIC 228 256
 FT CONFLICT 179 179
 FT CONFLICT 412 412
 SO SEQUENCE 503 AA; 54278 MW; AA6A39E0E272727C CRC64;

Query Match 10.1%; Score 66.5; DB 1; Length 503;
 Best Local Similarity 24.0%; Pred. No. 13;
 Matches 37; Conservative 16; Mismatches 52; Indels 49; Gaps 7;

OY 7 FLLMVGWGPVTEALIFETQXSLMAESENHLKTLGGCDADVR---GPP----- 53
 DB 105 FVVLKDNSSGDPADPHKVVYALSRCLKREBGTDTETAEARAAVPRPGGPFGLANTHA 164
 OY 54 -----GDSRLPAVOE-----WGAEPRVNLSPAIKHQFL-LTGD 86
 DB 165 GLDAPRLPARAGDEGLDLQAVQSCLDNHLTASWGA-DPVRTAARBGCGGLPLTGA 223
 OY 87 TQGRYCRSGSLTGMXQLSKLELT---GRKVL 117
 DB 224 CAGCGPLPAGELYGMA-----VETTPSPGHPAA 252

RESULT 11
 EPP1_HUMAN STANDARD: PRT; 5065 AA.
 ID EPP1_HUMAN
 AC P58107;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE EPIPLAKIN (450 KDA EPIDERMAL ANTIGEN).
 GN EPP1 OR EPIPL.
 OS Homo sapiens (Human).
 OS

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=21201183; PubMed=11278896;
 RA Fujiwara S., Takeo N., Ohtani Y., Parry D.A.D., Kunimatsu M., Lu R.,
 Sasaki M., Matsuo N., Khaleduzzaman M., Yoshioke H.,
 "Epilplakin, a novel member of the plakin family originally identified
 as a 450-kDa human epidermal autoantigen: structure and tissue
 localization.";
 RT J. Biol. Chem. 276:13340-13347(2001).
 RL TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN LIVER,
 CC SMALL INTESTINE, COLON, SALIVARY GLANDS, STOMACH AND APPENDIX.
 CC -1- SIMILARITY: BELONGS TO THE PLEKIN OR CYTOLINKER FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE PLEKIN OR CYTOLINKER FAMILY.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AB051895; BAB0803.1; -
 KM Colled coil; Repeat; Structural protein; Cytoskeleton.
 FT REPEAT 9 46 PLECTIN 1.
 FT REPEAT 47 84 PLECTIN 2.
 FT REPEAT 85 122 PLECTIN 3.
 FT REPEAT 123 160 PLECTIN 4.
 FT REPEAT 166 200 PLECTIN 5.
 FT REPEAT 253 290 PLECTIN 6.
 FT REPEAT 291 328 PLECTIN 7.
 FT REPEAT 330 366 PLECTIN 8.
 FT REPEAT 367 404 PLECTIN 9.
 FT REPEAT 521 558 PLECTIN 10.
 FT REPEAT 578 615 PLECTIN 11.
 FT REPEAT 616 653 PLECTIN 12.
 FT REPEAT 654 691 PLECTIN 13.
 FT REPEAT 692 729 PLECTIN 14.
 FT REPEAT 733 767 PLECTIN 15.
 FT REPEAT 840 878 PLECTIN 16.
 FT REPEAT 898 935 PLECTIN 17.
 FT REPEAT 936 973 PLECTIN 18.
 FT REPEAT 974 1011 PLECTIN 19.
 FT REPEAT 1012 1049 PLECTIN 20.
 FT REPEAT 1214 1251 PLECTIN 21.
 FT REPEAT 1252 1289 PLECTIN 22.
 FT REPEAT 1290 1327 PLECTIN 23.
 FT REPEAT 1328 1365 PLECTIN 24.
 FT REPEAT 1366 1403 PLECTIN 25.
 FT REPEAT 1539 1576 PLECTIN 26.
 FT REPEAT 1577 1614 PLECTIN 27.
 FT REPEAT 1615 1652 PLECTIN 28.
 FT REPEAT 1653 1690 PLECTIN 29.
 FT REPEAT 1694 1728 PLECTIN 30.
 FT REPEAT 1865 1902 PLECTIN 31.
 FT REPEAT 1903 1940 PLECTIN 32.
 FT REPEAT 1941 1978 PLECTIN 33.
 FT REPEAT 1979 2016 PLECTIN 34.
 FT REPEAT 2017 2054 PLECTIN 35.
 FT REPEAT 2192 2234 PLECTIN 36.
 FT REPEAT 2235 2272 PLECTIN 37.
 FT REPEAT 2273 2310 PLECTIN 38.
 FT REPEAT 2311 2348 PLECTIN 39.
 FT REPEAT 2352 2386 PLECTIN 40.
 FT REPEAT 2726 2768 PLECTIN 41.
 FT REPEAT 2769 2806 PLECTIN 42.
 FT REPEAT 2807 2844 PLECTIN 43.
 FT REPEAT 2845 2882 PLECTIN 44.
 FT REPEAT 2886 2920 PLECTIN 45.

FT REPEAT 3260 3302 PLECTIN 46.
 FT REPEAT 3303 3340 PLECTIN 47.
 FT REPEAT 3341 3378 PLECTIN 48.
 FT REPEAT 3379 3416 PLECTIN 49.
 FT REPEAT 3420 3454 PLECTIN 50.
 FT REPEAT 3494 3536 PLECTIN 51.
 FT REPEAT 3574 3616 PLECTIN 52.
 FT REPEAT 3637 3674 PLECTIN 53.
 FT REPEAT 3675 3712 PLECTIN 54.
 FT REPEAT 3754 3798 PLECTIN 55.
 FT REPEAT 4328 4370 PLECTIN 56.
 FT REPEAT 4371 4408 PLECTIN 57.
 FT REPEAT 4409 4446 PLECTIN 58.
 FT REPEAT 4447 4484 PLECTIN 59.
 FT REPEAT 4488 4522 PLECTIN 60.
 FT REPEAT 4522 4560 PLECTIN 61.
 FT REPEAT 4595 4632 PLECTIN 62.
 FT REPEAT 4682 4720 PLECTIN 63.
 FT REPEAT 4943 4980 PLECTIN 64.
 FT REPEAT 5022 5056 PLECTIN 65.
 FT DOMAIN 1935 1951 COILED COIL (POTENTIAL).
 FT DOMAIN 2593 2624 COILED COIL (POTENTIAL).
 FT DOMAIN 3127 3158 COILED COIL (POTENTIAL).
 FT DOMAIN 3661 3692 COILED COIL (POTENTIAL).
 FT DOMAIN 4195 4226 COILED COIL (POTENTIAL).
 FT DOMAIN 4729 4760 COILED COIL (POTENTIAL).
 SQ SEQUENCE 5065 AA; 553078 MW; 378A472DEDA19489 CRC64;
 Query Match 10.1%; Score 66.5; DB 1; Length 5065;
 Best Local Similarity 26.1%; Pred. No. 1; 9e+02;
 Matches 35; Conservative 11; Mismatches 45; Indels 43; Gaps 8;
 QY 9 LMGVYMGVPTFAIFETQXSLMAESENHLKTLQCCDADVGPQDSRLPAVQEN---- 64
 DB 1136 LLEDVQEGRTVPQLASVQR--WVQE---TKLAQARVMPGPRGE--VPVAV--WLDA 1186
 QY 65 -----GQEPFHL-DSPAIKHQFLITGDTGQRYCRSLSTGKQSLKLEL 110
 DB 1187 GIITQFTLEALAGTQSPQAVQAEQPAVKACLMGTG-----CVAGV-----LLOP 1230
 QY 111 TGPVYACSLADG 124
 DB 1231 SGAKASIAQAVRDG 1244
 RESULT 12
 ID 1B16_HUMAN STANDARD; PRT; 362 AA.
 AC P19373;
 DT 01-NOV-1990 (rel. 16, Created)
 DT 01-NOV-1990 (rel. 16, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-27 B*2703 ALPHA CHAIN
 DE PRECURSOR (B-27D).
 OS HLA-B OR HLAB.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=88227491; PubMed=3286582;
 RA Choo S.Y., St John T., Orr H.T., Hansen J.A.;
 RT "Molecular analysis of the variant alloantigen HLA-B*27D (HLA-B*2703)
 RT identifies a unique single amino acid substitution.";
 RL Hum. Immunol. 21:209-219(1988).
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
 THE IMMUNE SYSTEM.
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
 MICROGLOBULIN).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration


```

FT DISULFID 563 705 BY SIMILARITY.
FT DISULFID 572 667 BY SIMILARITY.
FT DISULFID 588 595 BY SIMILARITY.
FT DISULFID 734 758 BY SIMILARITY.
FT DISULFID 745 785 BY SIMILARITY.
FT DISULFID 763 765 BY SIMILARITY.
FT DISULFID 827 839 BY SIMILARITY.
FT DISULFID 823 843 BY SIMILARITY.
FT DISULFID 800 873 BY SIMILARITY.
FT DISULFID 870 873 BY SIMILARITY.
FT DISULFID 908 911 BY SIMILARITY.
FT DISULFID 942 1128 BY SIMILARITY.
FT DISULFID 1338 1339 BY SIMILARITY.
FT DISULFID 1355 1541 BY SIMILARITY.
FT DISULFID 1548 1573 BY SIMILARITY.
FT DISULFID 1568 1609 OR 1611 (BY SIMILARITY).
FT DISULFID 1641 1792 BY SIMILARITY.
FT DISULFID 1619 1754 BY SIMILARITY.
FT DISULFID 1596 1757 BY SIMILARITY.
FT DISULFID 1662 1670 BY SIMILARITY.
FT DISULFID 2393 2443 BY SIMILARITY.
FT DISULFID 2408 2457 BY SIMILARITY.
FT DISULFID 2419 2473 BY SIMILARITY.
FT DISULFID 2423 2475 BY SIMILARITY.
FT DISULFID ? 2480 BY SIMILARITY.
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 905 905 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1184 1184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1243 1243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1640 1640 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1892 1892 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1959 1959 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2026 2026 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2069 2069 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2215 2215 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2254 2254 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2459 2459 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 2482 AA: 272394 MW: 04987DDBFBCAEDD CRC64:

```

Query Match 10.0%; Score 66; DB 1; Length 2482;
 Best Local Similarity 33.3%; Pred. No. 94;
 Matches 16; Conservative 9; Mismatches 19; Indels 4; Gaps 1;

```

QY 61 VOEMAGQEPYHLDSPAKHQLTGTQGRYRRCRSGSTGKXQSLK 108
Db 1780 VOEMAVQOQPCQVPKQECQPVSGG---YOCQVLLSALFACCHKVL 1823

```

```

RESULT 14
BLAC_MYCTU STANDARD; PRT; 307 AA.
AC 010670;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BETA-LACTAMASE PRECURSOR (EC 3.5.2.6) (PENICILLINASE).
GN BLAA OR BLAC OR RV2068C OR MT2128 OR MTC149.07C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 201 / H37RA;
RC MEDLINE=97291283; PubMed=9145897;
RA Hackbart C.J., Unsal I., Chambers H.F.;
RT Cloning and sequence analysis of a class A beta-lactamase from
RT Mycobacterium tuberculosis H37Ra.
RL Anticrib. Agents Chemother. 41:1182-1185(1997).
RN 12;
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;

```

```

RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.
RL Nature 393:537-544(1998).
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gali J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: A BETA-LACTAM + H(2)O = A SUBSTITUTED BETA-
CC AMINO ACID.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67924; AAB07556.1; -
DR EMBL: 273966; CA98216.1; -
DR EMBL: AE007063; AAK46408.1; -
DR HSSP: P00808; IMBL.
DR TIGR: MT2128; -
DR InterPro: IP001466; Beta_lactam.
DR InterPro: IP000871; Beta_lactam_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE: PS00146; BETA-LACTAMASE_A; 1.
KW Hydrolase; Antibiotic resistance; Membrane; Lipoprotein; Signal;
KW Complete proteome.
FT STGNAL 1 23 POTENTIAL.
FT CHAIN 24 307 BETA-LACTAMASE.
FT LIPID 24 24 N-ACYL DIGLYCERIDE (POTENTIAL).
FT ACT SITE 84 84 BY SIMILARITY.
FT BINDING 250 252 SUBSTRATE (BY SIMILARITY).
SO SEQUENCE 307 AA: 32567 MW: 448CB2A0E05F4315 CRC64:

```

Query Match 9.8%; Score 65; DB 1; Length 307;
 Best Local Similarity 27.9%; Pred. No. 11;
 Matches 36; Conservative 17; Mismatches 44; Indels 32; Gaps 7;

```

QY 12 GVTWGPVTEAIAIYE--TOXSL-----WASESHXIKTLG---OCADVP----- 50
Db 128 GMTIGQLCDAAIRYSDGTANLLADGCPGCTAFTGYLSIGDVSRDLDEPELNR 187
QY 51 GPQDSR---LPVQEMGQEPYHLDSPAKHQLTGTG-----DTQGRYRRCRSGSTGKXQ 103
Db 188 DPQEDERTTPPAIALVALVQOVLVGNALPPDKRALLTDMARNRTGAKRIRAGPADW-- 245
QY 104 LSKLELTG 112
1:::11

```

Db 246 --KV1DKTG 252

RESULT 15

VP19_VZVVD STANDARD: PRT: 483 AA.

AC P09276;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 01-DEC-1992 (Rel. 24, Last annotation update)

DE CAPSID ASSEMBLY AND DNA MATURATION PROTEIN (CAPSID PROTEIN VP19C).

GN 20.

OS Varicella-zoster virus (strain Dumas) (VZV).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Varicellovirinae.

OX NCBI_Taxid:10338;

RP SEQUENCE FROM N.A.

RX MEDLINE:86306657; PubMed:3018124;

RA Davison A.J., Scott J.E.;

RT "The complete DNA sequence of varicella-zoster virus.";

RL J. Gen. Virol. 67:175-181(1986).

CC -1- FUNCTION: COMPONENT OF THE BASAL LAYER IN WHICH THE CAPSIDS ARE EMBEDDED. BINDS DNA (BY SIMILARITY).

CC -1- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP19C.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X04370; CAA27903.1; -.

DR PIR: B27343; W2B20.

DR Capsid assembly; Coat protein; DNA-binding.

KW SEQUENCE 483 AA; 53971 MW; A584CF73D689BF91 CRC64;

QY 56 SRPAAQWQKQEPVHLDSPAIKHOFILTDGQGRKRCGLSTGKXOLSK 106

Db 297 ANIPACVFWQVDRDLHLSADGLKHFVLYTGRORQREGVRLHLALSQLE 347

Query Match 9.8%; Score 65; DB 1; Length 483;

Best Local Similarity 31.4%; Pred. No. 18;

Matches 16; Conservative 9; Mismatches 26; Indels 0; Gaps 0;

RESULT 16

AFX1_HUMAN STANDARD: PRT: 504 AA.

AC P98177; Q13720; Q43821;

DT 01-OCT-1996 (Rel. 34, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PUTATIVE FORK HEAD DOMAIN TRANSCRIPTION FACTOR AFX1.

GN MLT7 OR AFX1 OR AFX.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_Taxid:9606;

RP SEQUENCE FROM N.A.

RC TISSUE:Blood;

MDLINE:98001080; PubMed:9341872;

RA Peters U., Haberland G., Kozrzewa M., Nolte D., Mueller U.;

RT AFX1 and p54nrb: fine mapping, genomic structure, and exclusion as candidate genes of X-linked dystonia parkinsonism.";

RL Hum. Genet. 100:569-572(1997).

RN 12;

RN SEQUENCE FROM N.A.

RP MEDLINE:97163401; PubMed:9010221;

RA Borkhardt A., Repp R., Haas O.A., Iels T., Harbott J., Kreuder J., Hammermann J., Henn T., Lampert F.;

RT "Cloning and characterization of AFX, the gene that fuses to MLL in acute leukemias with a t(x;11)(q13;q23).";

RL Oncogene 14:195-202(1997).

RN 13;

RP CHROMOSOMAL TRANSLOCATION.

RC TISSUE:Bone marrow;

RX MEDLINE:95118921; PubMed:7529552;

RA Parry P., Wei Y., Evans G.;

RT "Cloning and characterization of the t(x;11) breakpoint from a leukemic cell line identify a new member of the forkhead gene family.";

RL Genes Chromosomes Cancer 11:79-84(1994).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

CC -1- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND PANCREAS.

CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL TRANSLOCATION T(X;11)(Q13;Q23) THAT INVOLVES MLT7 AND MLL/HRX.

CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.

CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Y11284; CAA72156.1; -.

DR EMBL: Y11285; CAA72156.1; JOINED.

DR EMBL: Y11286; CAA72156.1; JOINED.

DR EMBL: X3396; CAA63819.1; -.

DR EMBL: U10072; AAA82171.1; ALT_SEQ.

DR MIM: 300033; -.

DR InterPro: IPR001766; Fork_head.

DR Pfam: PF00250; Fork_head; 1.

DR PRINTS: PR00053; FORKHEAD.

DR SMART: SM00339; FH; 1.

DR PROSITE: PS00657; FORK_HEAD_1; FALSE_NEG.

DR PROSITE: PS00658; FORK_HEAD_2; 1.

DR PROSITE: PS50039; FORK_HEAD_3; 1.

KW Transcription regulation; DNA-binding; Nuclear protein;

KW Chromosomal translocation; Proto-oncogene.

FT DNA_BIND 99 187

FT FT 1 33

FT FT 73 73

FT FT 78 78

FT FT 108 108

FT FT 421 421

SEQUENCE 504 AA; 53470 MW; 37767944FD32B8C4 CRC64;

QY 12 GYTWGPV--TEAAIFYETQXSLWA-----ESEHXLKTLGCGCDDVYGPQDSRLPVQEW 64

Db 332 GYV-GPLHYSSSLFPAEBEPLSAGGCGVSSQALAL--LTSDTTPPPADVMTQV--- 385

QY 65 GAQEPVHLDSPAIKHOFILTDGQGRKRCGLSTGKXOLSKILELTGPKVLAISLAL 122

Db 386 ---DPLISQAPTL---LLIGLPLS-----SSKLATGVLGCLPKPLLEAPGPSLVPPLISM 432

RESULT 17

ID RSFA_BACSU STANDARD: PRT: 258 AA.

AC P39650;

DT 01-FEB-1995 (Rel. 31, Created)


```

DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESPORE SPECIFIC TRANSCRIPTIONAL ACTIVATOR RSTA.
GN RSTA OR IPA-92R.
OS Bacillus subtilis.
OC Bacillaria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaeser P., Kunst F., Arnaut M., Coudart M.P., Gonzales M.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Prescan E., Sancha M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=20096685; PubMed=10629188;
RA Wu L.J., Errington J.;
RT "Identification and characterization of a new prespore-specific
RT regulatory gene, rsfA, of Bacillus subtilis.";
RL J. Bacteriol. 182:418-424(2000).
CC -!- FUNCTION: REGULATOR OF TRANSCRIPTION THAT FINE-TUNES GENE
CC EXPRESSION IN THE PRESPORE.
CC -!- SIMILARITY: TO B.SUBTILIS YLBO.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X73124; CAAS1648.1; -
DR EMBL: Z99123; CAB5789.1; -
DR PIR: S39747; S39747.
DR Subtilist: BG10638; rsfA.
KW Transcription regulation; Activator; DNA-binding; Complete proteome.
SQ SEQUENCE 258 AA; 29728 MW; 67C65A0483BFB5C CRC64;

Query Match 9.8%; Score 64.5; DB 1; Length 258;
Best Local Similarity 29.1%; Pred. No. 9.9;
Matches 25; Conservative 14; Mismatches 28; Indels 19; Gaps 6;

QY 12 GYWGCVTEAIFETQXSL-WASEHXLKTLGGCCADADVPGRGSRLL-PAV----- 61
DB 50 GFWNANVVRHQ--YERKALDLAKKORQRKALGN-----GGPARKRLLYOPAVDPETI 101
OY 62 QEWGAOEPRVHLDSPAIKH-OPLITGD 86
DB 102 QETAAEPVKTEPTSVENEOPLMSGE 127

RESULT 18
VIF_HV1MA STANDARD: PRT; 192 AA.
AC P04599;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION INFECTIVITY FACTOR (SOR PROTEIN).
GN VIF.
OS Human immunodeficiency virus type 1 (MAL isolate) (HIV-1).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11697;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=86245056; PubMed=2424612;
RA Allizon M., Main-Hobson S., Montagnier L., Sonigo P.;
RT "Genetic variability of the AIDS virus: nucleotide sequence analysis
RT of two isolates from African patients.";
RL Cell 46:63-74(1986).
CC -!- FUNCTION: PLIERMINES VIRUS INFECTIVITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X04415; CA828013.1; -
DR EMBL: A07116; CA00620.1; -
DR HIV: K03456; VIFSMAL.
DR InterPro: IPR000475; Viral_infect.
DR Pfam: PF00559; Vif; 1.
DR PRINTS: PR00349; VIRIONINFECT.
DR ProDom: PD000063; Viral_infect; 1.
KW AIDS.
SQ SEQUENCE 192 AA; 22723 MW; 1926C410DB92E255 CRC64;

Query Match 9.7%; Score 64; DB 1; Length 192;
Best Local Similarity 20.6%; Pred. No. 7.9;
Matches 21; Conservative 15; Mismatches 42; Indels 24; Gaps 3;

QY 8 LILMGV-----TWGVRFAAIFETQXSL-WASEHXLKTLGGCCADADVPGRGSRLLPAV 61
DB 8 MLYWQVDRMRKRTWISLVNHNMYGSKAKNRYRNHYESRHHKVSSEVNIPLGDRLVVR 67
OY 62 QEWGAOEPRVHLDSPAIKHOFLLTGDTGRCRSGSLSTGWQ 103
DB 68 TYWGLD-----TGERKD--WHIGHGVSIEMRQ 91

RESULT 19
S24B_ARATH STANDARD: PRT; 1069 AA.
AC S24B_ARATH
AC 09M081; O65535;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE PROTEIN TRANSPORT PROTEIN SEC24-LIKE AT4G32640.
GN AT4G32640 OR F4D11.160.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV_COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wandt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Sliemka W., Entlan K.-D., Terry N.,
RA Harris B., Ansoorge W., Brandt P., Griwell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
RA Kreis M., Delsen M., Pridmore P., Watson M., Schmidheini T.,
RA Reichert B., Fortetelle D., Perez-Alonso M., Bouty M., Bancroft I.,
RA Vos P., Honelisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzengerger T., Bothe G., Rameberger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., Van Staveren M., Dirse W.,
RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koelter P.,
RA Berner S., Hempel S., Feldpausch M., Lambers S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,

```

RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
 RA Petrecci A., Rajadran M.-A., Lyne M., Benes V., Reckmann S.,
 RA Borkova D., Bloedeker H., Scharte M., Grimm M., Loehnert T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Danner D., Herzl A.,
 RA Neumann S., Argilou A., Vitale D., Liguori R., Piravandi E.,
 RA Massenti O., Quigley F., Clabaud G., Muendlein A., Falber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lechary A., Aubourg S.,
 RA Chedron F., Cooke R., Berger C., Montfort A., Casachubeta E.,
 RA Gildons T., Weber N., Vandenbol M., Barges M., Tercel J., Torres A.,
 RA Perez-Ferez A., Furnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bleike C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedha N., Gnoj L., Cordes M., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoeckling T., Kallik J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Gelsel C., Layman D.,
 RA Du H., Ali J., Bernhoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 RA Chen E., Maria M., Martensen R., McCombie W.R.,
 RT *Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana.;
 RL Nature 402:769-777(1999).
 CC -1- FUNCTION: COMPONENT OF THE COP1 COAT, THAT COVERS ER-DERIVED
 CC VESICLES INVOLVED IN TRANSPORT FROM THE ENDOPLASMIC RETICULUM TO
 CC THE GOLGI APPARATUS. COP1 IS COMPOSED OF AT LEAST FIVE PROTEINS:
 CC THE SEC23/24 COMPLEX, THE SEC13/31 COMPLEX, AND THE PROTEIN SAR1.
 CC ACTS IN THE CYTOSOL TO PROMOTE THE TRANSPORT OF SECRETORY,
 CC PLASMA MEMBRANE, AND VACUOLAR PROTEINS FROM THE ENDOPLASMIC
 CC RETICULUM TO THE GOLGI COMPLEX (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SEC23/SEC24 FAMILY. SEC24 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----
 CC DR EMBL: AL022537; CA18597.1; ALT-SEQ.
 DR EMBL: AL161581; CAB79981.1; -
 KW Hypothetical protein; Transport; Protein transport; Golgi stack;
 KW Endoplasmic reticulum; Multigene family.
 FT DOMAIN 430 455 ZINC-FINGER-LIKE.
 FT DOMAIN 55 60
 FT DOMAIN 341 344 POLY-SER.
 FT DOMAIN 358 361 POLY-PRO.
 SQ SEQUENCE 1069 AA; 114775 MW; 81CD4290B9FA3FD7 CRC64;

Query Match 9.7% Score 64; DB 1: Length 1069;
 Best Local Similarity 29.5% Pred. No. 58;
 Matches 26; Conservative 8; Mismatches 34; Indels 20; Gaps 4;

OY 42 LGCCADVAGP-----PDSSRLPAVOEMGAQEPVHLSDPAIKHQFL-TCPTQGR 90
 DB 318 KQCPATVYGFPSRIDPNQIPRGSSSPVYFETROSQANPPRPATSDYVKTGNCSPR 377
 OY 91 Y-CRSG-----LSTGWXLDSLKLE 109
 DB 378 YMRCTINQIPCTVDLSTSGMDLALMVQ 405

RESULT 20
 BUB1_HUMAN

ID BUB1_HUMAN STANDARD; PRT; 1085 AA.
 AC 043683; 060626; 043643; 043430;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MITOTIC CHECKPOINT SERINE/THREONINE-PROTEIN KINASE BUB1 (EC 2.7.1.-)
 DE (HUBB1) (BUB1A).
 GN BUB1 OR BUB1L.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. AND VARIANT TYR-492.
 RX MEDLINE:98180623; PubMed:9521327;
 RA Cahill D.P., Lengauer C., Yu J., Klagsins G.J., Willson J.K.V.,
 RA Markowitz S.D., Kinzler K.W., Vogelstein B.;
 RT "Mutations of mitotic checkpoint genes in human cancers.";
 RL Nature 392:300-303(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:99005160; PubMed:9790499;
 RA Ouyang B., Lan Z., Meadows J., Pan H., Fukasawa K., Li W., Dai W.;
 RT "Human Bubi: a putative spindle checkpoint kinase closely linked to
 RT cell proliferation.";
 RL Cell Growth Differ. 9:877-885(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE:Testis;
 RA Seeley T.W.;
 RL submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:98327111; PubMed:9660858;
 RA Taylor S.S., Ha E., McKoon F.;
 RT "The human homologue of Bub3 is required for kinetochore localization
 RT of Bub1 and a Mad3/Bub1-related protein kinase.";
 RL J. Cell Biol. 142:1-11(1998).
 RN [5]
 RP SEQUENCE FROM N.A. AND VARIANTS ASP-36 AND ARG-648.
 RX MEDLINE:99296833; PubMed:10366450;
 RA Cahill D.P., da Costa L.T., Carson-Walter E.B., Kinzler K.W.,
 RA Vogelstein B., Lengauer C.;
 RT "Characterization of MAD2B and other mitotic spindle checkpoint
 RT genes.";
 RL Genomics 58:181-187(1999).
 RN [6]
 RP SEQUENCE OF 276-1085 FROM N.A.
 RX MEDLINE:98110573; PubMed:9441741;
 RA Pangilinan F., Li Q., Weaver T., Lewis B.C., Dang C.V., Spencer F.;
 RT "Mammalian BUB1 protein kinases: map positions and in vivo
 RT expression.";
 RL Genomics 46:379-388(1997).
 CC -1- FUNCTION: INVOLVED IN CELL CYCLE CHECKPOINT ENFORCEMENT. CAN
 CC INTERACT AND PHOSPHORYLATE BUB3.
 CC -1- ENZYME REGULATION: AUTOPHOSPHORYLATED WHEN THE CELLS ENTERS
 CC MITOSIS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR IN INTERPHASE CELLS. KINETOCHORE
 CC LOCALIZATION IS REQUIRED FOR NORMAL MITOTIC TIMING AND CHECKPOINT
 CC RESPONSE TO SPINDLE DAMAGE (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: HIGH EXPRESSION IN TESTIS AND THYMUS. LESS IN
 CC COLON, SPLEEN, LUNG AND SMALL INTESTINE. EXPRESSED IN FETAL
 CC THYMUS, BONE MARROW, HEART, LIVER, SPLEEN AND THYMUS. EXPRESSION
 CC IS ASSOCIATED WITH CELLS/TISSUES WITH A HIGH MITOTIC INDEX.
 CC -1- INDUCTION: INHIBITED BY PHORBOL 12-MYRISTATE 13-ACETATE (PMA).
 CC -1- DOMAIN: CD1 DOMAIN DIRECTS KINETOCHORE LOCALIZATION AND BINDING TO
 CC BUB3.
 CC -1- DISEASE: DEFECTS IN BUB1 ARE ASSOCIATED WITH TUMOR FORMATION.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC BUB1 SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 CD1 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: AF011387; AAC39546.1; -;
 CC EMBL: AF043294; AAB97855.2; -;
 CC EMBL: AF045078; AAC12729.1; -;
 CC EMBL: AF047471; AAC03122.1; -;
 CC EMBL: AF053305; AAC06259.1; -;
 CC EMBL: AF133363; AAC43675.1; -;
 CC EMBL: AF133349; AAC43675.1; JOINED.
 CC EMBL: AF133350; AAC43675.1; JOINED.
 CC EMBL: AF133351; AAC43675.1; JOINED.
 CC EMBL: AF133352; AAC43675.1; JOINED.
 CC EMBL: AF133353; AAC43675.1; JOINED.
 CC EMBL: AF133354; AAC43675.1; JOINED.
 CC EMBL: AF133355; AAC43675.1; JOINED.
 CC EMBL: AF133356; AAC43675.1; JOINED.
 CC EMBL: AF133357; AAC43675.1; JOINED.
 CC EMBL: AF133358; AAC43675.1; JOINED.
 CC EMBL: AF133359; AAC43675.1; JOINED.
 CC EMBL: AF133360; AAC43675.1; JOINED.
 CC EMBL: AF133361; AAC43675.1; JOINED.
 CC EMBL: AF133362; AAC43675.1; JOINED.
 CC MIM: 602452; -;
 CC DR InterPro: IPR000719; Euk_pkinase.
 CC DR InterPro: IPR002290; Ser_thr_kin_actsite.
 CC DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
 CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 CC DR Transferrase: Serine/threonine-protein kinase: ATP-binding; Cell cycle;
 CC KW Nuclear protein; Mitosis; Phosphorylation; Polymorphism.
 CC FT DOMAIN 11 160
 CC FT NP_BIND 787 1085
 CC FT BINDING 821 801
 CC FT ACT_SITE 917 821
 CC FT DOMAIN 58 65
 CC FT VARIANT 36 36
 CC FT VARIANT 492 492
 CC FT VARIANT 648 648
 CC FT CONFLICT 70 70
 CC FT CONFLICT 276 279
 CC SQ SEQUENCE 1085 AA: 122375 MW: 38A5E51F86C53BDC CRC64:

Query Match 9.7%; Score 64; DB 1; Length 1085;
 Best Local Similarity 21.4%; Pred. No. 59;
 Matches 34; Conservative 19; Mismatches 52; Indels 54; Gaps 7;

OY 10 LKGV-----TWGPTVAITFTQ-XSLMAESEHXLK-----TLGQ 44
 DB 581 VMGIRCKNTLAPSPKSGDPTSAOLAISFPFHLKLPVESVHLEEDKENVAKOCQTQATLDS 640
 OY 45 CDADVQGPDCSLRPAVQEWGAGQEP-----VHLDSPAIKHQHLLLGDNQ-GHYRC 93
 DB 641 CEENWVVPSPKGFSPLOEKSPPQALSSHWYSASLLRSLPAGG--VLTCAEALGVEAC 698
 OY 94 R-----SGLTGCMXQLSKLLETGPKVL 116
 DB 699 RLITDTDAIAIEDPPDAIAGLGAEWMQMSISGTYDAPNFI 737

RESULT 21
 CAL4_HUMAN
 ID CAL4_HUMAN STANDARD; PRT; 1669 AA.
 AC P02462;

DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
 GN COL4A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89340433; PubMed=2701944;
 RA Solinen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;
 RT "Structural organization of the gene for the alpha 1 chain of human
 RL type IV collagen.";
 RL J. Biol. Chem. 264:13565-13571(1989).
 RN [2]
 RP SEQUENCE OF 46-1257 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88083584; PubMed=3691802;
 RA Solinen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
 RT "Complete primary structure of the alpha 1-chain of human basement
 RL membrane (type IV) collagen.";
 RL FEBS Lett. 225:188-194(1987).
 RN [3]
 RP SEQUENCE OF 1-943 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88029471; PubMed=3311751;
 RA Brazel D., Oberbaumer I., Dieringer H., Babel W., Glanville R.W.,
 RA Deutzmann R., Kuehn K.;
 RT "Completion of the amino acid sequence of the alpha 1 chain of human
 RT basement membrane collagen (type IV) reveals 21 non-triple
 RT interruptions located within the collagenous domain.";
 RL Eur. J. Biochem. 168:529-536(1987).
 RN [4]
 RP SEQUENCE OF 28-243.
 RX MEDLINE=86004708; PubMed=4043082;
 RA Glanville R.W., Qian R.Q., Stebold B., Ristell J., Kuehn K.;
 RT "Amino acid sequence of the N-terminal aggregation and cross-linking
 RT region (75 domain) of the alpha 1 (IV) chain of human basement
 RL membrane collagen.";
 RL Eur. J. Biochem. 152:213-219(1985).
 RN [5]
 RP SEQUENCE OF 534-1447.
 RX MEDLINE=85003629; PubMed=6434307;
 RA Babel W., Glanville R.W.;
 RT "Structure of human basement-membrane (type IV) collagen. Complete
 RT amino-acid sequence of a 914-residue-long pepsin fragment from the
 RL alpha 1(IV) chain.";
 RL Eur. J. Biochem. 143:545-556(1984).
 RN [6]
 RP SEQUENCE OF 1256-1669 FROM N.A.
 RX MEDLINE=85207819; PubMed=2581969;
 RA Phlajantien T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
 RA Cheung M.-C., Prockop D.J., Boyd C.D.;
 RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
 RT procollagen reveal an unusual homology of amino acid sequences in two
 RT halves of the carboxyl-terminal domain.";
 RL J. Biol. Chem. 260:7681-7687(1985).
 RN [7]
 RP SEQUENCE OF 1259-1669 FROM N.A.
 RX MEDLINE=85216555; PubMed=2582422;
 RA Brinker J.M., Gudes L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
 RA Kefalides N.A., Myers J.C.;
 RT "Restricted homology between human alpha 1 type IV and other
 RT procollagen chains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
 RN [8]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=89034231; PubMed=3182844;
 RA Solinen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
 RT collagen are divergently encoded on opposite DNA strands and have an

RT overlapping promoter region.";
 RL J. Biol. Chem. 263:17217-17220(1988).
 RN [9]
 RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
 RC TISSUE-Placenta;
 RX MEDLINE:89005112; PubMed:2844531;
 RA Siebold B., Deutzmann R., Kuehn K.;
 RT "The arrangement of intra- and intermolecular disulfide bonds in the
 RT carboxyterminal, non-collagenous aggregation and cross-linking domain
 RT of basement-membrane type IV collagen.";
 RL Eur. J. Biochem. 176:617-624(1988).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOCFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: LYSINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT ARE HYDROXYLATED IN ALL CASES AND BIND CARBOHYDRATE.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 CC EMBL: M26576; AAA53098.1; -
 CC DR EMBL: J04217; AAA53098.1; JOINED.
 CC DR EMBL: M26550; AAA53098.1; JOINED.
 CC DR EMBL: M26540; AAA53098.1; JOINED.
 CC DR EMBL: M26542; AAA53098.1; JOINED.
 CC DR EMBL: M26543; AAA53098.1; JOINED.
 CC DR EMBL: M26544; AAA53098.1; JOINED.
 CC DR EMBL: M26545; AAA53098.1; JOINED.
 CC DR EMBL: M26346; AAA53098.1; JOINED.
 CC DR EMBL: M26347; AAA53098.1; JOINED.
 CC DR EMBL: M26537; AAA53098.1; JOINED.
 CC DR EMBL: M26538; AAA53098.1; JOINED.
 CC DR EMBL: M26548; AAA53098.1; JOINED.
 CC DR EMBL: M26549; AAA53098.1; JOINED.
 CC DR EMBL: M26551; AAA53098.1; JOINED.
 CC DR EMBL: M26552; AAA53098.1; JOINED.
 CC DR EMBL: M26553; AAA53098.1; JOINED.
 CC DR EMBL: M26554; AAA53098.1; JOINED.
 CC DR EMBL: M26555; AAA53098.1; JOINED.
 CC DR EMBL: M26556; AAA53098.1; JOINED.
 CC DR EMBL: M26357; AAA53098.1; JOINED.
 CC DR EMBL: M26359; AAA53098.1; JOINED.
 CC DR EMBL: M26558; AAA53098.1; JOINED.
 CC DR EMBL: M26559; AAA53098.1; JOINED.
 CC DR EMBL: M26560; AAA53098.1; JOINED.
 CC DR EMBL: M26561; AAA53098.1; JOINED.
 CC DR EMBL: M26562; AAA53098.1; JOINED.
 CC DR EMBL: M26536; AAA53098.1; JOINED.
 CC DR EMBL: M26563; AAA53098.1; JOINED.
 CC DR EMBL: M26541; AAA53098.1; JOINED.
 CC DR EMBL: M26564; AAA53098.1; JOINED.
 CC DR EMBL: M26565; AAA53098.1; JOINED.
 CC DR EMBL: M26566; AAA53098.1; JOINED.

DR EMBL: M26567; AAA53098.1; JOINED.
 DR EMBL: M26568; AAA53098.1; JOINED.
 DR EMBL: M26569; AAA53098.1; JOINED.
 DR EMBL: M26570; AAA53098.1; JOINED.
 DR EMBL: M26571; AAA53098.1; JOINED.
 DR EMBL: M26572; AAA53098.1; JOINED.
 DR EMBL: M26573; AAA53098.1; JOINED.
 DR EMBL: M26574; AAA53098.1; JOINED.
 DR EMBL: M26575; AAA53098.1; JOINED.
 DR EMBL: Y00706; CAA68698.1; -
 DR EMBL: X05561; CAA29075.1; -
 DR EMBL: M10940; AAA52006.1; -
 DR EMBL: M11315; AAA52042.1; -
 DR PIR: S16876; CGH04B.
 DR MIM: 120130; -
 DR InterPro: IPR001442; C4.
 DR InterPro: IPR000887; Collagen.
 DR Pfam: PF01413; C4; 2.
 DR Pfam: PF01391; Collagen; 21.
 DR ProDom: PD003923; C4; 2.
 DR SMART: SM00111; C4; 2.
 KW Extracellular matrix; Connective tissue; Basement membrane;
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 27
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
 FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .).
 FT DISULFID 1460 1551 OR 1548.
 FT DISULFID 1493 1548 OR 1551.
 FT DISULFID 1505 1511 OR 1551.
 FT DISULFID 1570 1665 OR 1662.
 FT DISULFID 1604 1662 OR 1665.
 FT DISULFID 1616 1622 OR 1662.
 FT CONFLICT 237 238 SG -> KE (IN REF. 4).
 FT CONFLICT 241 241 G -> K (IN REF. 4).
 FT CONFLICT 319 319 Q -> A (IN REF. 3).
 FT CONFLICT 719 719 N -> D (IN REF. 5).
 FT CONFLICT 837 837 D -> Y (IN REF. 5).
 FT CONFLICT 842 842 K -> P (IN REF. 5).
 FT CONFLICT 896 896 V -> W (IN REF. 2).
 FT CONFLICT 904 904 E -> Q (IN REF. 5).
 FT CONFLICT 914 914 S -> K (IN REF. 5).
 FT CONFLICT 998 998 S -> K (IN REF. 5).
 FT CONFLICT 1010 1030 K -> P (IN REF. 5).
 FT CONFLICT 1012 1032 S -> K (IN REF. 5).
 FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
 SQ SEQUENCE 1669 AA; 3BEBA6DFB9B8A84 CRC64;
 Query Match 9.6%; Score 63.5; DB 1; Length 1669;
 Best Local Similarity 36.7%; Pred. No. 1,1e+02;
 Matches 18; Conservative 7; Mismatches 15; Indels 9; Gaps 3;
 QY 50 PGPPGDSRLPAVQEWGAEFVHDSPAIKHQFLTDTQ--GRRGRSG 96
 DB 1422 PGPPGDRGLP-----GSMGPR--GTPSVYDGLFLVTHSQITDDPGQPSG 1463
 RESULT 22
 ID 1C12_HUMAN STANDARD: PRT; 366 AA.
 AC P30505;
 DT 01-APR-1993 (rel. 25, Created)
 DT 01-APR-1993 (rel. 25, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, CW-8 CW*0801 ALPHA CHAIN
 DE PRECURSOR (CW8.1).
 GN HLA-C OR HLA-C.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

```

OX NCBI_TaxId-9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93031775; PubMed-1384166;
RA Zemmour J., Gumperz J.E., Hildebrand W.H., Ward F.E., Marsh S.G.,
RA Williams R.C., Parham P.;
RT "The molecular basis for reactivity of anti-Cw1 and anti-Cw3
RT allantoisera with HLA-B*6 haplotypes.";
RL Tissue Antigens 39:249-257(1992).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M84174; AAA59688.1; -
DR HSSP: P30460; IAGB.
DR MIM: 142840; -
DR InterPro: IPR003006; Iq_MHC.
DR InterPro: IPR003597; Iq_C1.
DR InterPro: IPR001039; MHC_I.
DR Pfam: PF00047; Iq_1.
DR Pfam: PF00129; MHC_I; 1.
DR PRODOM: PD000050; MHC_I; 1.
DR SMART: SM00407; IG_C1; 1.
DR PROSITE: PS00290; IG_MHC; 1.
DR MHC_I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 1 25 366
FT FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT FT CW-8 CW*0801 ALPHA CHAIN.
FT DOMAIN 115 114 EXTRACELLULAR ALPHA-1.
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.
FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.
FT TRANSMEM 309 333 CONNECTING PEPTIDE.
FT DOMAIN 334 366 CYTOPLASMIC TAIL.
FT DISULFID 125 188 BY SIMILARITY.
FT DISULFID 227 283 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 110 110
SQ SEQUENCE 366 AA; 40772 MW; 2A84D41389A0486A CRC64;

Query Match 9.5%; Score 63; DB 1; Length 366;
Best Local Similarity 24.8%; Pred. No. 21;
Matches 34; Conservative 14; Mismatches 67; Indels 22; Gaps 5;

OY 1 MSMLVFLLLMGV-----TWGPTVEAIFETQXSLAEESEHLKTLG-----QCDA 47
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 4 MAPRTLILLSGALALTEWTACSHSMRYF-TAVSRGRGPERPIAVGYDDTQFOFQDS 62
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 48 DVGPPGDSRLPAVOEKGAEQVPHLSDPAIKHFLLTGTQGRRCRSGSLSTGMXOLSKL 107
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 63 DAASPRGEPAPVWEDEG---PEYWDRETQKYKRAQTDRVSLRNLR-----GYNNOSEA 114
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 108 LELTGPVYLACSLALDG 124
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 115 GSHTLQRMGCDLGPDG 131

RESULT 23
ID 1C13 HUMAN STANDARD: PRT; 366 AA.
AC P30506;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

```

```

DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, CW-8 CW*0802 ALPHA CHAIN
DE PRECURSOR (CW8.2).
GN HLA-C OR HLAC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId-9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93031775; PubMed-1384166;
RA Zemmour J., Gumperz J.E., Hildebrand W.H., Ward F.E., Marsh S.G.,
RA Williams R.C., Parham P.;
RT "The molecular basis for reactivity of anti-Cw1 and anti-Cw3
RT allantoisera with HLA-B*6 haplotypes.";
RL Tissue Antigens 39:249-257(1992).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M84173; AAA59687.1; -
DR HSSP: P30460; IAGB.
DR MIM: 142840; -
DR InterPro: IPR003006; Iq_MHC.
DR InterPro: IPR003597; Iq_C1.
DR InterPro: IPR001039; MHC_I.
DR Pfam: PF00047; Iq_1.
DR Pfam: PF00129; MHC_I; 1.
DR PRODOM: PD000050; MHC_I; 1.
DR SMART: SM00407; IG_C1; 1.
DR PROSITE: PS00290; IG_MHC; 1.
DR MHC_I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 1 25 366
FT FT HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT FT CW-8 CW*0802 ALPHA CHAIN.
FT DOMAIN 115 114 EXTRACELLULAR ALPHA-1.
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.
FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.
FT TRANSMEM 309 333 CONNECTING PEPTIDE.
FT DOMAIN 334 366 CYTOPLASMIC TAIL.
FT DISULFID 125 188 BY SIMILARITY.
FT DISULFID 227 283 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 110 110
SQ SEQUENCE 366 AA; 40871 MW; D343B054568EA52C CRC64;

Query Match 9.5%; Score 63; DB 1; Length 366;
Best Local Similarity 24.8%; Pred. No. 21;
Matches 34; Conservative 14; Mismatches 67; Indels 22; Gaps 5;

OY 1 MSMLVFLLLMGV-----TWGPTVEAIFETQXSLAEESEHLKTLG-----QCDA 47
| : | | | | | | | | | | | | | | | | | | | | | |
DB 4 MAPRTLILLSGALALTEWTACSHSMRYF-TAVSRGRGPERPIAVGYDDTQFOFQDS 62
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 48 DVGPPGDSRLPAVOEKGAEQVPHLSDPAIKHFLLTGTQGRRCRSGSLSTGMXOLSKL 107
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 63 DAASPRGEPAPVWEDEG---PEYWDRETQKYKRAQTDRVSLRNLR-----GYNNOSEA 114
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 108 LELTGPVYLACSLALDG 124
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 115 GSHTLQRMGCDLGPDG 131

RESULT 24

```



```

DB 291 RDNMTGTNRPILVISPDLISRVGYLCAGLPDTPRGVADQFTGSGTSPMONGYGVKGFG 350
QY 94 -----RSG-----LSTGWQSLKLELTGPRVACSLADGAS 126
DB 351 FROGSDVMGRTISRTSRSGFELLRIKNGWTQTSK--EDEVGRVYVDNLMWGSYS 403

RESULT 26
PARC_ECOLI STANDARD; PRT; 752 AA.
ID PARC_ECOLI
AC P20082;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TOPOISOMERASE IV SUBUNIT A (EC 5.99.1.7).
GN PARC OR B3019 OR Z4373 OR EC53903.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-91004247; PubMed-2170028;
RA Kato J.-I., Nishimura Y., Imanura R., Niki H., Hiraga S., Suzuki H.;
RT "New topoisomerase essential for chromosome segregation in E. coli.";
RL Cell 63:393-404(1990).
RN [2]
RP ERRATUM.
RA Kato J.-I., Nishimura Y., Imanura R., Niki H., Hiraga S., Suzuki H.;
RL Cell 65:1289-1290(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE-21074935; PubMed-11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE-21156331; PubMed-11258796;
RA Hayashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 724-752 FROM N.A.
RX MEDLINE-92212294; PubMed-1557036;
RA Coleman J.;
RT "Characterization of the Escherichia coli gene for
RT 1-acyl-sn-glycerol-3-phosphate acyltransferase (PlsC).";
RL Mol. Gen. Genet. 232:295-303(1992).
RN [7]

```

```

RP REVISIONS, AND CHARACTERIZATION.
RC STRAIN-K12;
RX MEDLINE-94043292; PubMed-8227000;
RA Peng H., Mariani K.J.;
RT "Escherichia coli topoisomerase IV. Purification, characterization,
RT subunit structure, and subunit interactions.";
RL J. Biol. Chem. 268:24481-24481(1993).
CC -!- FUNCTION: TOPOISOMERASE IV IS ESSENTIAL FOR CHROMOSOME
CC SEGREGATION. IT HAS RELAXATION OF SUPERCOILED DNA ACTIVITY.
CC PERFORMS THE DECATENATION EVENTS REQUIRED DURING THE REPLICATION
CC OF A CIRCULAR DNA MOLECULE.
CC -!- SUBUNIT: COMPOSED OF TWO SUBUNITS: PARC AND PARE.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC -!- SIMILARITY: STRONG, WITH THE A SUBUNIT OF GYRASE.
CC -----
CC This SWISS-PROT entry is copyright: It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M58408; AAA24297.1; ALT_INIT.
DR EMBL: M63491; AAA24396.1; -
DR EMBL: U28377; AAA69187.1; -
DR EMBL: AE000384; AAC76055.1; -
DR EMBL: AE005531; AAC58155.1; -
DR EMBL: AP002563; BAB37326.1; -
DR EMBL: L22025; AAC36840.1; -
DR PIR: A36075; A36075.
DR PIR: A39936; A39936.
DR Ecogene; EG10686; PARC.
DR InterPro: IPR002205; DNA_topoisoy.
DR Pfam; PF00521; DNA_topoisoy.1.
DR SMART; SM00434; TOPAc; 1.
KW Topoisomerase; Isomerase; DNA-binding; Complete proteome.
FT ACT_SITE 120 120 DNA_CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 752 AA; 83831 MW; 0D4907E96CEE7086 CRC64;

Query Match 9.5%; Score 63; DB 1; Length 752;
Best Local Similarity 34.1%; Pred. No. 49;
Matches 31; Conservative 5; Mismatches 33; Indels 22; Gaps 5;

QY 11 MGYTWGPVTEALIFETQXSLMASEHXLTGQCDAD-VPGPGDSRUPAVQEWGAQGP 69
DB 106 WCAPDDPKSPFAMRY-TESRLSKYSFELLSELGGQADWVPNDGT-----LQEP 154
QY 70 VHLDSPAIKHQFLTGDPTGGRYCRSGLSYG 100
DB 155 KML-PARLPNLLNGT-----GIANG 175

RESULT 27
GUXB_CELFT STANDARD; PRT; 1090 AA.
ID GUXB_CELFT
AC P50899;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE EXOGLUCANASE B PRECURSOR (EC 3.2.1.91) (EXOCELLULOBIHYDROLASE B)
DE (1,4-BETA-CELLULOBIHYDROLASE B) (CBP120).
GN CBHB OR CENE.
OS Cellulomonas fimi.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
RC STRAIN-ATCC 484;
RX MEDLINE-96003898; PubMed-7575482;
RA Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;

```

```

RT "Cellulohydrolase B, a second exo-cellulohydrolase from the
RT cellulolytic bacterium Cellulomonas flm1."
RT Biochem. J. 311:67-74(1995).
RN [2]
RX MEDLINE:93209933; PubMed:8458833;
RA Melnik A., Gilles N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT "Cellulose-binding polypeptides from Cellulomonas flm1: endoglucanase
RT D (Cend), a family A beta-1,4-glucanase."
RT J. Bacteriol. 175:1910-1918(1993).
RN [3]
RP SEQUENCE OF 54-78.
RX MEDLINE:94197708; PubMed:8147863;
RA Shen H., Tome P., Melnik A., Gilles N.R., Kilburn D.G.,
RA Warren R.A.J., Miller R.C. Jr.;
RT "Stereochemical course of hydrolysis catalysed by Cellulomonas flm1
RT Cend, a member of a new family of beta-1,4-glucanases."
RL Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
CC -1- FUNCTION: HYDROLYSES CELLULOSE TO A MIXTURE OF CELLOTETRAOSE,
CC CELLOTRIOSE AND CELLOBIOSE, WITH ONLY A TRACE OF GLUCOSE. IT
CC HYDROLYSES CELLULOSE TO CELLOTRIOSE AND CELLOBIOSE, AND
CC CELLOTETRAOSE TO CELLOBIOSE, BUT IT DID NOT HYDROLYSE CELLULOSE.
CC HAS ALSO WEAK ENDOGLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS
CC WITH INVERSION OF ANOMERIC CONFIGURATION.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES
CC IN CELLULOSE AND CELLOTETRAOSE, RELEASING CELLOBIOSE FROM THE NON-
CC REDUCING ENDS OF THE CHAINS.
CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY L (FAMILY 48 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L38827; AAB00822.1; -.
DR HSSP: P07986; 1EXH.
DR INTERPRO: IPR001919; CBD_2.
DR INTERPRO: IPR001777; FN_III.
DR INTERPRO: IPR000556; Glyco_hydro_48.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF00041; fn3; 3.
DR Pfam: PF02011; Glyco_hydro_48; 1.
DR PRINTS: PR00014; FNTYPEIII.
DR PRINTS: PR00844; GLHYDRLASE48.
DR PRODOM: PD011903; Glyco_hydro_48; 1.
DR SMART: SM00060; FN3; 3.
DR PROSITE: PS00361; CBD_BACTERIAL; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
KW FT SIGNAL 1 33
FT PROPEP 34 53
FT CHAIN 54 1090
FT DOMAIN 54 689
FT DOMAIN 700 785
FT DOMAIN 794 884
FT DOMAIN 891 978
FT DOMAIN 989 1090
FT ACT_SITE 513 513
FT DISULFID 990 1089
FT SEQUENCE 1090 AA; 114829 MW; 046BB9D956F2F399 CRC64;

```

```

Query Match 9.5%; Score 63; DB 1; Length 1090;
Best Local Similarity 17.6%; Pred No. 76;
Matches 33; Conservative 19; Mismatches 44; Indels 92; Gaps 5;
7 FILLGCTWGPVTE-----AAIFETQXSLMAESENHLKT----- 41

```

```

DB 353 YLLSYTMAMGATDTSSGAMRIGSSHAHFGYONPLAAMALSTDPKLTSPKADMAA 412
OY 42 -----LGGCDADVPGPEDSRU 58
DB 413 SMORLEPTWLOASNGCIAGCATNSWDCAVYAPPACTPFYCMGYTEAPVYVDPSSNM 472
OY 59 PAVOENGAEVPIHDPALKHOFLLTGDPGRYRCRSGSLTGWXSLSLELTGPKVLAC 118
DB 473 FCGMANGVGRVAMEL-----YYASGNAQAK-----KILDKVPPVVA- 508
OY 119 STALDGAS 126
DB 509 NISTDGAS 516
RESULT 28
CP2B_HUMAN STANDARD; PRT; 508 AA.
ID CP2B_HUMAN
AC 015528;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 25-HYDROXYVITAMIN D-1 ALPHA-HYDROXYLASE, MITOCHONDRIAL PRECURSOR
DE (EC 1.14.-.-) (25-OH-D-1 ALPHA-HYDROXYLASE) (P450C1 ALPHA) (P450VD1-
DE ALPHA-HYDROXYLASE) (VD3 1A HYDROXYLASE)
GN CYP27B1 OR CYP27B OR CYP1ALPHA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:98089075; PubMed:9428799;
RA Fu G.K., Portale A.P., Miller W.L.;
RT "Complete structure of the human gene for the vitamin D 1alpha-
RT hydroxylase, P450c1alpha."
RL DNA Cell Biol. 16:1499-1507(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE: Kidney;
RX MEDLINE:98008873; PubMed:9344864;
RA Monkawa T., Yoshida T., Wakino S., Shinkai T., Anezawa H., Deluca H.F.,
RA Suda T., Hayashi M., Saruta T.;
RT "Molecular cloning of cDNA and genomic DNA for human
RT 25-hydroxyvitamin D3 1 alpha-hydroxylase."
RL Biochem. Biophys. Res. Commun. 239:527-533(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE:98075882; PubMed:9415400;
RA Fu G.K., Lin D., Zhang Y.H., Bikle D.D., Shackleton C.H., Miller W.L.,
RA Portale A.A.;
RT "Cloning of human 25-hydroxyvitamin D-1 alpha-hydroxylase and
RT mutations causing vitamin D-dependent rickets type 1."
RL Mol. Endocrinol. 11:1961-1970(1997).
CC -1- FUNCTION: CATALYZES THE CONVERSION OF 25-HYDROXYVITAMIN D3
CC (25(OH)D) TO 1-ALPHA,25-DIHYDROXYVITAMIN D3 (1,25(OH)2D) PLAYS AN
CC IMPORTANT ROLE IN CRUCIAL ROLE IN NORMAL BONE GROWTH, CALCIUM
CC METABOLISM, AND TISSUE DIFFERENTIATION.
CC -1- PATHWAY: SECOND STEP IN THE CONVERSION OF VITAMIN D(3) INTO THE
CC ACTIVE FORM (1-ALPHA,25-DIHYDROXYVITAMIN D(3)).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- TISSUE SPECIFICITY: KIDNEY.
CC -1- DISEASE: DEFECTS IN CYP27B ARE A CAUSE OF VITAMIN D-DEPENDENT
CC RICKETS TYPE 1 (VDDR-1), A DISEASE CHARACTERIZED BY MUSCLE
CC WEAKNESS AND RICKETS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



```

CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL: J03758: AAA37439.1: -
DR EMBL: M23333: AAA51625.1: -
DR EMBL: J04694: AAA50292.1: -
DR EMBL: X06777: CAA29946.1: -
DR EMBL: X02201: CAA26132.1: -
DR EMBL: M15832: AAA37340.1: -
DR EMBL: M14042: AAA37342.1: -
DR EMBL: M12879: AAA37343.1: -
DR EMBL: M13024: -; NOT_ANNOTATED_CDS.
DR EMBL: M13025: -; NOT_ANNOTATED_CDS.
DR EMBL: M13026: AAA37344.1: -
DR EMBL: M13027: AAA37345.1: -
DR EMBL: M13043: AAA37346.1: -
DR EMBL: J04448: AAA37437.1: -
DR PIR: A33525: CSM84B.
DR MGI: M8454: Col4a1.
DR InterPro: IPR001442: C4.
DR InterPro: IPR000087: Collagen.
DR Pfam: PF01413: C4; 2.
DR Pfam: PF01391: Collagen; 21.
DR ProDom: PD003923: C4; 2.
DR SMART: SM00111: C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
Repeat; Hydroxylation; Glycoprotein; Collagen; signal.
FT SIGNAL 1 27
FT PROPEP 28 1672 AMINO-TERMINAL PROPEPTIDE (75 DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IIV) CHAIN.
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
FT DISULFID 1505 1511 BY SIMILARITY.
FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
FT DISULFID 1616 1622 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 26 26 A -> P (IN REF. 2).
FT CONFLICT 186 186 S -> L (IN REF. 2).
FT CONFLICT 319 319 Q -> S (IN REF. 2).
FT CONFLICT 369 369 Q -> L (IN REF. 2).
FT CONFLICT 403 403 L -> F (IN REF. 2).
FT CONFLICT 481 481 Q -> L (IN REF. 2).
FT CONFLICT 493 493 Q -> H (IN REF. 2).
FT CONFLICT 712 712 S -> I (IN REF. 2).
FT CONFLICT 813 813 E -> Q (IN REF. 2).
FT CONFLICT 982 982 Q -> H (IN REF. 2).
FT CONFLICT 1397 1397 V -> S (IN REF. 3).
SQ SEQUENCE 1669 AA: 160680 MW: 42916B91E5205BE9 CRC64:

Query Match 9.5%; Score 62.5; DB 1; Length 1669;
Best Local Similarity 38.5%; Pred. No. 1.4e+02;
Matches 15; Conservative 6; Mismatches 11; Indels 7; Gaps 2;
OY 50 PCPPCDRLPAVQEMGAEPVHLDSPAIKHOFLLTGDQ 88
DB 1422 PCPPGPDGLP-----GSMGPP--CTPSVDHGLVTRHSQ 1453

RESULT 30
1B25_HUMAN
ID 1B25_HUMAN STANDARD: PRT: 362 AA.
AC P30471:
DT 01-APR-1993 (rel. 25, Created)
DT 01-APR-1993 (rel. 25, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-35 B*3505 ALPHA CHAIN
DE PRECURSOR (B35-C).
CN HLA-B OR HLAB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Granlata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
[1]
RX MEDLINE:9226955; PubMed:1317015;
RA Bellch M.P., Madrigal J.A., Hildebrand W.H., Zemmour J.,
RA Williams R.C., Luz R., Petzl-Erler M.L., Parham P.;
RT "Unusual HLA-B alleles in two tribes of Brazilian Indians.";
RL Nature 357:326-329(1992).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL: M84385: AAA59635.1: -
DR HSSP: P30685: IAIN.
DR MIM: 142830: -
DR InterPro: IPR003006: I9_MHC.
DR InterPro: IPR003597: I9_C1.
DR InterPro: IPR001039: MHC_I.
DR Pfam: PF00047: I9_1.
DR Pfam: PF00129: MHC_I_1.
DR ProDom: PD000050: MHC_I_1.
DR SMART: SM00407: I0C1; 1.
DR PROSITE: PS00290: I9_MHC; 1.
KW MHC I; Transmembrane; Glycoprotein; signal.
FT SIGNAL 1 24
FT CHAIN 25 362
FT FT DOMAIN 25 114 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT FT DOMAIN 115 206 B-35 B*3505 ALPHA CHAIN.
FT FT DOMAIN 207 298 EXTRACELLULAR ALPHA-1.
FT FT DOMAIN 299 308 EXTRACELLULAR ALPHA-2.
FT FT DOMAIN 309 332 CONNECTING PEPTIDE.
FT FT TRANSMEM 333 362 CYTOPLASMIC TAIL.
FT FT DOMAIN 110 110 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT FT CARBOHYD 125 188 BY SIMILARITY.
FT FT DISULFID 227 283 BY SIMILARITY.
FT FT DISULFID 227 283 BY SIMILARITY.
SQ SEQUENCE 362 AA: 40374 MW: 9E601F49238BC37 CRC64:

Query Match 9.4%; Score 62; DB 1; Length 362;
Best Local Similarity 25.0%; Pred. No. 27;
Matches 33; Conservative 15; Mismatches 62; Indels 22; Gaps 6;
OY 6 VFLLMGV-----TWGPVTEALIFETQXSLMAESHXLKTLG-----QCDAVDPG 52
DB 9 VILLILGAVALTFTWAGSHSMRYFY-TAMSPRGCPRPFLAVGVYDQTFVFRPDSASP 67
OY 53 PGDSRLPAVQEMGAEPVHLDSPAIKHOFLLTGDQGRYRCRSGLSTGXKXSKLLLELG 112
DB 68 RTEPRAPWIEQGS--PEYWD---RNTQIFKTNTO-TYRESLRNLRGYNOSEAGSHTL 119

OY 113 PKVLACSLALDG 124
DB 120 QSMYGCGLGPDG 131

RESULT 31
1C04_HUMAN
ID 1C04_HUMAN STANDARD: PRT: 366 AA.
AC P30502:
DT 01-APR-1993 (rel. 25, Created)
DT 01-APR-1993 (rel. 25, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)

```

```

DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, CW-2 CW*0202 ALPHA CHAIN
DE PRECURSOR (CW2.2).
GN HLA-C OR HLAC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=89235215; PubMed=2715640;
RA Parham P., Lawlor D.A., Lomen C.E., Ennis P.D.;
RT "Diversity and diversification of HLA-A,B,C alleles.";
RL J. Immunol. 142:3937-3950(1989).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M24030; AAA59671.1; -
DR HSSP: P30685; IAIN.
DR MIM: 142840; -
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003597; Ig_C1.
DR InterPro: IPR001039; MHC_I.
DR Pfam: PF00047; Ig_1.
DR Pfam: PF00129; MHC_I; 1.
DR ProDom: PD000050; MHC_I; 1.
DR SMART: SM00407; IgC1; 1.
DR PROSITE: PS00290; IG_MHC; 1.
DR MHC I; Transmembrane; Glycoprotein; signal.
KM CHAIN 1 34
FT 25 366
FT 25 366 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 25 114 CW-2 CW*0202 ALPHA CHAIN.
FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.
FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.
FT DOMAIN 309 333 CONNECTING PEPTIDE.
FT TRANSMEM 309 333
FT DOMAIN 334 366 CYTOPLASMIC TAIL.
FT DISULFID 125 188 BY SIMILARITY.
FT DISULFID 227 283 BY SIMILARITY.
FT CARBOHYD 110 110 N-LINKED (GLCNAC... ) (BY SIMILARITY).
SQ SEQUENCE 366 AA; 40994 MW; 64AD7B8C2BCD5D6B CRC64;

Query Match 9.4%; Score 62; DB 1; Length 366;
Best Local Similarity 24.1%; Pred. No. 27;
Matches 33; Conservative 17; Mismatches 65; Indels 22; Gaps 5;

```

```

AC P30508;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, CW*1201 ALPHA CHAIN PRECURSOR
DE (HLA-CX52).
GN HLA-C OR HLAC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88330144; PubMed=2843461;
RA Taketa H., Inoko H., Ando A., Haranaka M., Watanabe B., Tsuji K.,
RA Iri H.;
RT "Cloning and analysis of HLA class I cDNA encoding a new HLA-C
RT specificity CX52.";
RL Immunogenetics 28:265-270(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA Taketa H., Sonoda A., Beck S., Heyes J.M., Bodmer J.G., Inoko H.;
RA Subtilted (JUN-1992) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M21963; AAA59847.1; -
DR HSSP: P30685; IAIN.
DR MIM: 142840; -
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003597; Ig_C1.
DR InterPro: IPR001039; MHC_I.
DR Pfam: PF00047; Ig_1.
DR Pfam: PF00129; MHC_I; 1.
DR ProDom: PD000050; MHC_I; 1.
DR SMART: SM00407; IgC1; 1.
DR PROSITE: PS00290; IG_MHC; 1.
DR MHC I; Transmembrane; Glycoprotein; signal.
KM CHAIN 1 24
FT 25 366
FT 25 366 HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT DOMAIN 25 114 CW*1201 ALPHA CHAIN.
FT DOMAIN 115 206 EXTRACELLULAR ALPHA-1.
FT DOMAIN 207 298 EXTRACELLULAR ALPHA-2.
FT DOMAIN 299 308 EXTRACELLULAR ALPHA-3.
FT DOMAIN 309 333 CONNECTING PEPTIDE.
FT TRANSMEM 309 333
FT DOMAIN 334 366 CYTOPLASMIC TAIL.
FT DISULFID 125 188 BY SIMILARITY.
FT DISULFID 227 283 BY SIMILARITY.
FT CARBOHYD 110 110 N-LINKED (GLCNAC... ) (BY SIMILARITY).
SQ SEQUENCE 366 AA; 40851 MW; E6A7DA16117A7B46 CRC64;

Query Match 9.4%; Score 62; DB 1; Length 366;
Best Local Similarity 24.1%; Pred. No. 27;
Matches 33; Conservative 15; Mismatches 67; Indels 22; Gaps 5;

```

```

Db 63 DAASPRGEPRAPWEOEG---PEYMDRETQKYYKQAQADRSLSRLNR-----GYNQSEA 114
OY 108 LELTGPVLAAGSLADG 124
Db 115 GSHTLQRMVCGDLCPDG 131

RESULT 33
CG95_HUMAN STANDARD: PRT: 620 AA.
AC 008379;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GOLIN-95.
GN GOLGA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE: Liver;
RA MEDLINE:93301617; PubMed:8315394;
RA Fritzier M.J., Hamel J.C., Ochs R.L., Chan E.K.L.;
RT "Molecular characterization of two human autoantigens: unique cDNAs
RT encoding 95- and 160-kD proteins of a putative family in the Golgi
RT complex.";
RL J. Exp. Med. 178:49-62(1993).
CC -1- FUNCTION: GOLGI AUTO-ANTIGEN; MAY HAVE A FUNCTION IN THE
CC PROCESSING AND TRANSPORT OF PROTEINS THROUGH THE GOLGI.
CC -1- DOMAIN: EXTENDED ROD-LIKE PROTEIN WITH COILED-COIL DOMAINS.
CC -1- SIMILARITY: HIGH, TO RAT CIS-GOLGI MATRIX PROTEIN GM130.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L06147; AAA35920.1; -.
CC DR MIM: 602580; -.
CC KM Golgi stack; Coiled coil.
CC FT DOMAIN 5 336 COILED COIL (POTENTIAL).
CC FT DOMAIN 352 510 COILED COIL (POTENTIAL).
CC FT DOMAIN 67 73 POLY-PRO.
CC FT DOMAIN 322 333 POLY-GLU.
CC FT SEQUENCE 620 AA: 70472 MW: 54B31A0FBB42BFC8 CRC64;
SQ

Query Match 9.4%; Score 62; DB 1; Length 620;
Best local Similarity 36.2%; Pred. No. 50;
Matches 21; Conservative 2; Mismatches 19; Indels 16; Gaps 2;

OY 26 EFGXSLMAESEXHLKTL---GQCDDADVPQPDGSRPLPAVQ-----MCAQ 67
Db 78 EFGQQLAEAEHRLKEELAGQLAQVQDNEGLSRNREGERLLELEAAELMGEQ 135

RESULT 34
Y029_HUMAN STANDARD: PRT: 971 AA.
AC 015032;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0029.
GN KIAA0029.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE: Bone marrow;
RX MEDLINE:96051387; PubMed:7584026;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawabayashi Y.,
RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid
RT cell line KG-1.";
RL DNA Res. 1:27-35(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D21852; BAA04878.1; -.
CC DR Interpro: IPR001374; R3H.
CC DR Pfam: PF01424; R3H; 1.
CC DR SMART: SM00393; R3H; 1.
CC RT Hypothetical protein.
CC FT DOMAIN 461 480 POLY-PRO.
CC FT SEQUENCE 971 AA: 107673 MW: D07684D368955108 CRC64;
SQ

Query Match 9.4%; Score 62; DB 1; Length 971;
Best local Similarity 29.3%; Pred. No. 85;
Matches 22; Conservative 11; Mismatches 24; Indels 18; Gaps 4;

OY 49 VQPPGDSRLPAVQENCAEPVHDSPAIRHOFILTGDTQ-----GRYCRSGLSL- 99
Db 794 VPG-QGDSRYPLL-----GQPLQYNPPAVLHGHIHPQCGPGRSGHNGRRQAKKAASFD 847
OY 100 ---GMXQLSKLELFT 111
Db 848 LGAGETVGVKLELT 862

```

```

RESULT 35
ACAL_ARATH STANDARD: PRT: 1020 AA.
ID ACAL_ARATH
AC Q37145; Q37146; Q42571; Q42587; Q9SFY1;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CALCIUM-TRANSPORTING ATPASE 1, PLASMA MEMBRANE-TYPE (EC 3.8.3.8)
DE (CA2+-ATPASE, ISOFORM 1) (PLASTID ENVELOPE ATPASE 1).
GN ACAL OR PEAL OR AT1G27770 OR T22C5.23 OR P28U5.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eudicots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP PRELIMINARY SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN: CV. COLUMBIA;
RX MEDLINE:94052104; PubMed:8234257;
RA Huang L., Berkelman T., Franklin A.E., Hoffman N.E.;
RT "Characterization of a gene encoding a Ca(2+)-ATPase-like protein in
RT the plastid envelope.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10066-10070(1993).
RN [2]
RP REVISIONS.
RA Huang L., Berkelman T., Franklin A.E., Hoffman N.E.;
RL Proc. Natl. Acad. Sci. U.S.A. 91:9664-9664(1994).
RN [3]
RP SEQUENCE FROM N.A.

```

RC STRAIN-CV. COLUMBIA;
 RA MEDLINE-21016719; PubMed-11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
 RA Dunn P., Eguia P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT *Sequence and analysis of chromosome 1 of the plant *Arabidopsis*
 RT *thaliana*.";
 RL Nature 408:816-820(2000).
 CC -1- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE HYDROLYSIS
 CC OF ATP COUPLED WITH THE TRANSLLOCATION OF CALCIUM FROM THE CYTOSOL
 CC OUT OF THE CELL OR INTO ORGANELLES.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + CA(2+)(CIS) = ADP + PHOSPHATE +
 CC CA(2+)(TRANS)
 CC -1- ENZYME REGULATION: ACTIVATED BY CALMODULIN (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.. CHLOROPLAST
 CC ENVELOPE (INNER MEMBRANE) (PROBABLE).
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHER LEVELS IN ROOTS THAN IN
 CC LEAVES.
 CC -1- DOMAIN: THE N-TERMINUS CONTAINS AN AUTOINHIBITORY CALMODULIN-
 CC BINDING DOMAIN, WHICH BINDS CALMODULIN IN A CALCIUM-DEPENDENT
 CC FASHION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (EI-E2 ATPASES). SUBFAMILY IIB.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO WRONG EXON
 CC BOUNDARIES PREDICTED FROM THE GENOMIC SEQUENCE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 CC EMBL: L08468; AAD10211.1; -
 CC EMBL: L08469; AAD10212.1; -
 CC EMBL: D13983; BAA03090.1; -
 CC EMBL: D13984; BAA03091.1; -
 CC EMBL: X69940; CAA49558.1; -
 CC EMBL: X69941; CAA49559.1; -
 CC EMBL: AC012375; AAF24958.1; ALT_SEQ.
 CC EMBL: AC079280; AAG50579.1; -
 CC HSSP: P11719; 1EUL.
 CC -----
 CC Mendel: 16033; Arabid. 2439; 16033.
 CC Mendel: 16034; Arabid. 2439; 16034.
 CC Mendel: 22293; Arabid. 2439; 22293.
 CC Mendel: 22294; Arabid. 2439; 22294.
 CC Mendel: 36857; Arabid. 2439; 36857.
 CC Mendel: 36858; Arabid. 2439; 36858.
 CC InterPro: IPR001757; EI-E2_ATPase.
 CC InterPro: IPR000695; HATPase.
 CC InterPro: IPR001454; Hydrolase.
 CC InterPro: IPR000661; Na_H_K_ATPase.
 CC Pfam: PF00122; EI-E2_ATPase; 1.
 CC Pfam: PF00702; Hydrolase; 1.
 CC PRINTS: PR00119; CATATPASE.
 CC PRINTS: PR00120; HATPASE.
 CC PROSITE: PS00154; ATPASE_EI_E2; 1.
 CC Hydrolase; Calcium transport; Transmembrane; Phosphorylation;
 CC ATP-binding; Metal-binding; Magnesium; Chloroplast;
 CC

KW Calmodulin-binding; Multigene family.
 FT DOMAIN 1 162 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 163 183 POTENTIAL.
 FT DOMAIN 184 201 LUMENAL (POTENTIAL).
 FT TRANSMEM 202 222 POTENTIAL.
 FT DOMAIN 223 350 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 351 370 POTENTIAL.
 FT DOMAIN 371 400 LUMENAL (POTENTIAL).
 FT TRANSMEM 401 418 POTENTIAL.
 FT DOMAIN 419 813 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 814 832 POTENTIAL.
 FT DOMAIN 833 843 LUMENAL (POTENTIAL).
 FT TRANSMEM 844 864 POTENTIAL.
 FT DOMAIN 865 884 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 885 907 POTENTIAL.
 FT DOMAIN 908 919 LUMENAL (POTENTIAL).
 FT TRANSMEM 920 941 POTENTIAL.
 FT DOMAIN 942 959 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 960 981 POTENTIAL.
 FT DOMAIN 982 991 LUMENAL (POTENTIAL).
 FT TRANSMEM 992 1013 POTENTIAL.
 FT DOMAIN 1014 1020 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 21 32 CALMODULIN-BINDING (BY SIMILARITY).
 FT MOD_RES 46 46 PHOSPHORYLATION (BY CDPK) (BY
 FT MOD_RES 456 456 SIMILARITY).
 FT MOD_RES 758 758 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 762 762 MAGNESIUM (BY SIMILARITY).
 FT CONFLICT 88 88 P -> S (IN REF. 1: AAD10211 AND REF. 2:
 FT CONFLICT 801 801 BAA03090 AND CAA49559).
 FT CONFLICT 801 801 I -> T (IN REF. 3 AND REF. 2: BAA03091
 FT AND CAA49558).
 SQ SEQUENCE 1020 AA: 11273 MW: 1894DESCIAAFRAF CRC64;
 Query Match 9.4%; Score 62; DB 1; Length 1020;
 Best local Similarity 28.4%; Pred. No. 89;
 Matches 19; Conservative 10; Mismatches 30; Indels 8; Gaps 2;
 QY 65 GAGEPVHDSPAIKHOFLLTGDTGGRGRS-----GLSTGWXLSKLELTGPKVLACS 119
 DB 291 GGESEPV---SVSEHDFLLSGTFKVGQSCMLTWTYGMRTQWCKLMWATLSEGDDETPIQ 347
 QY 120 LALDGAS 126
 DB 348 VKLNGVA 354
 RESULT 36
 ALA7_ARATH STANDARD: PRT: 1247 AA.
 ID ALA7_ARATH
 AC Q91VK9;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE 7 (EC 3.6.3.13).
 GN ALA7 OR AT3G13900 OR WDC16.2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
 OC eurosids II: Brassicales: Brassicaceae: Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RX MEDLINE-20277480; PubMed-10819329;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.;
 RT *Structural analysis of Arabidopsis thaliana chromosome 3 I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
 RT clones.";
 RL DNA Res. 7:131-135(2000).
 RN [2]
 RP CONCEPTUAL TRANSLATION.

RA Axelisen K.B. ;
 RL Unpublished observations (NOV-2000).
 CC -1- FUNCTION: INVOLVED IN TRANSPORT OF PHOSPHOLIPIDS (POTENTIAL).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O -- ADP + PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 (E1-E2 ATPASES). SUBFAMILY IV.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO WRONG EXON
 BORDERS. PREDICTED FROM THE GENOMIC SEQUENCE.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

 CC
 DR EMBL: AB019229; BAB02320.1; ALT_SEQ.
 DR InterPro: IPR001757; E1-E2_ATPase.
 DR InterPro: IPR001454; Hydrolase.
 DR Pfam: PF00702; Hydrolase: 1.
 DR PROSITE: PS00154; ATPASE_E1_E2: 1.
 KM Hydrolase; Transmembrane; Phosphorylation; ATP-binding; Metal-binding;
 KM Magnesium; Multigene family; Hypothetical protein.
 FT DOMAIN 1
 FT TRANSMEM 75 96
 FT DOMAIN 97 100
 FT TRANSMEM 101 123
 FT DOMAIN 124 305
 FT TRANSMEM 306 327
 FT DOMAIN 328 359
 FT TRANSMEM 360 377
 FT DOMAIN 378 945
 FT TRANSMEM 946 965
 FT DOMAIN 966 979
 FT TRANSMEM 980 999
 FT DOMAIN 1000 1029
 FT TRANSMEM 1030 1052
 FT DOMAIN 1053 1065
 FT TRANSMEM 1066 1088
 FT DOMAIN 1089 1094
 FT TRANSMEM 1095 1115
 FT DOMAIN 1116 1132
 FT TRANSMEM 1133 1157
 FT DOMAIN 1158 1247
 FT MOD_RES 425 425
 FT METAL 880 880
 FT METAL 894 894
 SQ SEQUENCE 1247 AA; 140928 MW; 07E323DB71B6D9FB CRC64;

OS Didelphis marsupialis virginiana (North American opossum).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Didelphis.
 OX NCBI_TaxID:9267;
 RN [1]
 RP SEQUENCE, AND MASS SPECTROMETRY.
 RC TISSUE=Serum;
 RA Neves-Ferreira A.G.C., Perales J., Fox J.W., Shannon J.D.,
 RA Makino D.L., Garrett R.C., Domont G.B.;
 RT "Structural and functional analyses of DMA3, a snake venom
 RT metalloproteinase inhibitor from Didelphis marsupialis serum.";
 RL Submitted (FEB-2001) to the SWISS-PROT data bank.
 CC -1- FUNCTION: METALLOPROTEINASE INHIBITOR.
 CC -1- TISSUE SPECIFICITY: BLOOD AND MILK.
 CC -1- PTM: ALL THE "X" ARE PROBABLY GLYCOSYLATED ASPARAGINE.
 CC -1- MASS SPECTROMETRY: MW:42691; METHOD=MALDI.
 CC -1- SIMILARITY: TO HUMAN ALPHA-1B-GLYCOPROTEIN.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 DR InterPro: IPR003006; I9_MHC.
 DR InterPro: IPR003599; I9.
 DR InterPro: IPR003600; I9-like.
 DR Pfam: PF00047; I9; 1.
 DR SMART: SM00409; I6; 2.
 DR SMART: SM00410; I6-like; 1.
 KM Metalloproteinase inhibitor; Venom; Glycoprotein;
 KM Immunoglobulin domain.
 FT DOMAIN 1
 FT TRANSMEM 21 81
 FT DOMAIN 81 272
 FT DISULFID 28 74
 FT DISULFID 121 163
 FT DISULFID 213 265
 SQ SEQUENCE 291 AA; 32374 MW; 1B89D62F1A2828BD CRC64;

Query Match 9.4%; Score 62; DB 1; Length 1247;
 Best Local Similarity 23.6%; Pred. No. 11e+02;
 Matches 26; Conservative 14; Mismatches 40; Indels 30; Gaps 3;
 OY 26 ETQXSLMAESENKLTLCQCDADVPQPCDSRLPAVQEWCAQPVHLDSPKIHQPLNG 85
 DB 692 ETEYSITWNEPFIHAKTSVCADR-----EMLEKVDMMKELIIVG 732
 QY 86 DTQGRYRCRSGSLTGXQSL-----KILELTGPKV-----LACSLALDG 124
 DB 733 AVAVEDKLOKGVQCIDKLAQGLKIMVLTDGDMETAINIGVACSLRQG 782
 RESULT 37
 DM43_DIDMA STANDARD; PRT; 291 AA.
 AC P82957;
 DT 20-AUG-2001 (Rel. 40; Created)
 DT 20-AUG-2001 (Rel. 40; Last sequence update)
 DT 20-AUG-2001 (Rel. 40; Last annotation update)
 DE VENOM METALLOPROTEINASE INHIBITOR DMA3.

Query Match 9.3%; Score 61.5; DB 1; Length 291;
 Best Local Similarity 42.9%; Pred. No. 24;
 Matches 15; Conservative 5; Mismatches 14; Indels 1; Gaps 1;
 OY 83 LTQDGRYRCRSGSL-TGMXQSLKLELTGPKVL 116
 DB 64 VTADNRGIYRCRLITSENDWTPLSAPVEVYGRPL 98
 RESULT 38
 EVAL_MOUSE STANDARD; PRT; 591 AA.
 AC P97767; O08818;
 DT 15-JUL-1998 (Rel. 36; Created)
 DT 15-JUL-1998 (Rel. 36; Last sequence update)
 DT 20-AUG-2001 (Rel. 40; Last annotation update)
 DE EYES ABSENT HOMOLOG 1.
 CN EVAL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID:10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Embryo.
 RA MEDLINE-97158722; PubMed-9006082;
 RA Xu P.-X., Woo I., Her H., Beier D.R., Maas R.L.;
 RT "Mouse Eye homologues of the Drosophila eyes absent gene require Pax6
 RT for expression in lens and nasal placode.";
 RL Development 124:219-231(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC STRAIN=CB/20;
 RA MEDLINE-97172972; PubMed-9020840;
 RA Abdelhak S., Kalatzis V., Hellig R., Compain S., Samson D.,
 RA Vincent C., Weil D., Graud C., Sahly I., Jethovici M.,
 RA Bitner-Glindzicz M., Francis M., Lacombe D., Vigneron J.,
 RA Charachon R., Boven K., Bebeder P., van Regemorter N.,
 RA Welschbach J., Petit C.;


```
FT CHAIN 174 276 32 KDA ENAMELIN.
FT CHAIN 515 665 25 KDA ENAMELIN.
FT CHAIN 670 7 34 KDA ENAMELIN.
FT CHAIN 7 7 45 KDA ENAMELIN.
FT MOD_RES 53 53 53 PHOSPHORYLATION (PROBABLE).
FT MOD_RES 191 191 191 PHOSPHORYLATION.
FT MOD_RES 216 216 216 PHOSPHORYLATION.
FT MOD_RES 547 547 547 HYDROXYLATION.
FT CARBOHYD 245 245 245 N-LINKED (GLCNAC. . .).
FT CARBOHYD 252 252 252 N-LINKED (GLCNAC. . .).
FT CARBOHYD 264 264 264 N-LINKED (GLCNAC. . .).
FT CARBOHYD 291 291 291 N-LINKED (GLCNAC. . .).
FT CARBOHYD 462 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 929 929 929 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1040 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 680 680 680 H -> D (IN REF. 2).
FT CONFLICT 838 840 840 H -> T (IN REF. 2).
SQ SEQUENCE 1142 AA; 128352 MW; 938306BC87CC5FC6 CRC64;
```

```
Query Match
Best Local Similarity 19.7%; Score 61.5; DB 1; Length 1142;
Matches 14; Conservative 9; Mismatches 29; Indels 19; Gaps 1;
```

```
OY 28 QXSLMAESEXHLTLGQ-----CDADVPGGPDSRLPAVQEWCAOE 68
ID VIF_HV1A2 STANDARD; PRT; 3591 AA.
AC P03402;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION INFECTIVITY FACTOR (SOR PROTEIN).
```

```
OY 69 PVHLDSPAIFKH 79
ID VIF_HV1A2 STANDARD; PRT; 3591 AA.
AC P03402;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE FILAMENTOUS HEMAGGLUTININ.
```

```
DB 670 LSFKEDEPTVRH 680
```

```
RESULT 40
FHAB_BORPE STANDARD; PRT; 3591 AA.
AC P12355;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE FILAMENTOUS HEMAGGLUTININ.
GN FHAB.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID:520;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE:90355839; PubMed:2388559;
RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Genetic characterization of Bordetella pertussis filamentous
haemagglutinin: a protein processed from an unusually large
precursor."
RL Mol. Microbiol. 4:787-800(1990).
RN [2]
RP SEQUENCE OF 1-3261 FROM N.A.
RX MEDLINE:89202384; PubMed:2539596;
RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide
sequence and crucial role in adherence."
RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
CC -1- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND
INFECTION.
CC -1- SUBCELLULAR LOCATION: SURFACE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
```

```
DR EMBL: M60351; AAA22974.1; -.
DR EMBL: M60351; AAA22975.1; ALT_INIT.
DR EMBL: M60351; AAA22976.1; ALT_INIT.
KW Antigen; Hemagglutinin.
SQ SEQUENCE 3591 AA; 367420 MW; EF7418B30D65138 CRC64;
```

```
Query Match
Best Local Similarity 30.6%; Score 61.5; DB 1; Length 3591;
Matches 22; Conservative 13; Mismatches 28; Indels 9; Gaps 3;
```

```
OY 26 EFXSLMAESEXHLK--TLGQCDADVPGGPDSRLPA----VQEWG---AOEVLHDSFA 76
ID VIF_HV1A2 STANDARD; PRT; 192 AA.
AC P03402;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION INFECTIVITY FACTOR (SOR PROTEIN).
```

```
OY 77 IKHOFLLTGDTQ 88
ID VIF_HV1A2 STANDARD; PRT; 192 AA.
AC P03402;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION INFECTIVITY FACTOR (SOR PROTEIN).
```

```
DB 1800 IENTAKLSGEVQ 1811
```

```
RESULT 41
VIF_HV1A2 STANDARD; PRT; 192 AA.
ID VIF_HV1A2 STANDARD; PRT; 192 AA.
AC P03402;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION INFECTIVITY FACTOR (SOR PROTEIN).
```

```
GN VIF.
OS Human immunodeficiency virus type 1 (HIV-1) (HIV-1).
OC Viruses; Retrovirus; Retroviridae; Lentivirus.
OX NCBI_TaxID:11685;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:85090453; PubMed:2578227;
RA Sanchez-Pescador R., Power M.D., Barr P.J., Steimer K.S.,
RA Stemple M.M., Brown-Shimer S.L., Gee W.W., Renard A., Randolph A.,
RA Levy J.A., Dina D., Lucif P.A.;
RT "Nucleotide sequence and expression of an AIDS-associated retrovirus
(RARV-2)."
RL Science 227:484-492(1985).
```

```
CC -1- FUNCTION: DETERMINES VIRUS INFECTIVITY.
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
```

```
DR EMBL: K02007; AAB59877.1; -.
DR PIR: A04003; ASLJ01.
DR HIV: K02007; VIFSS52.
DR InterPro: IPR000475; Viral_infect.
DR Pfam: PF00559; Vif_1.
DR PRINTS: PR00349; VIRIONINFECT.
DR ProDom: PD000063; Viral_infect; 1.
KW AIDS.
SQ SEQUENCE 192 AA; 22460 MW; 0E993D33FF1134EB CRC64;
```

```
Query Match
Best Local Similarity 9.2%; Score 61; DB 1; Length 192;
Matches 19; Conservative 18; Mismatches 39; Indels 24; Gaps 3;
```

```
OY 8 LLLMGV-----TWGPTENAIFETQXSLMAESEXHLKTLGQCDADVPGGPDSRLPAV 61
ID VIF_HV1A2 STANDARD; PRT; 192 AA.
AC P03402;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION INFECTIVITY FACTOR (SOR PROTEIN).
```

```
DB 8 MIMQVDRMIRPMKGLVKNHMYISKAKCMFYRNHNESTHPRVSEVNHPLDADKLVIT 67
```

```
OY 62 QEWCAOEVLHDSFAIFKH 101
ID VIF_HV1A2 STANDARD; PRT; 192 AA.
AC P03402;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE VIRION INFECTIVITY FACTOR (SOR PROTEIN).
```

```
DB 68 TYWG-----LHTGERE--WHLGQVAIEW 89
```


CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL109732; CAB52045.1; -
 DR InterPro: IPR000073; Abhydrolase.
 DR InterPro: IPR000379; Est_Lip_thioest_actsite.
 DR InterPro: IPR002410; Pro_aminoptase.
 DR Pfam: PF00561; abhydrolase; 1.
 DR PRINTS: PR00793; PROAMNOPTASE.
 KW Hydrolase; Amino peptidase.
 FT ACT_SITE 114 114 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 271 271 BY SIMILARITY.
 FT ACT_SITE 299 299 PROTON DONOR (BY SIMILARITY).
 SQ SEQUENCE 323 AA: 35040 MW: 7BEF0133B353EEB CRC64;

Query Match 9.2%: Score 61; DB 1; Length 323;
 Best Local Similarity 25.0%; Pred. No. 30;
 Matches 23; Conservative 13; Mismatches 32; Indels 24; Gaps 4;

OY 9 LMGVWGPV-----TEAIFVE---TQXSLMASEHKLKLGCD-----AD 48
 DB 109 LVWGVSWGVGLGRVQTHPGVTELVGVATGSNAEVALITRGIGITFPEAHERFLAE 168
 OY 49 VPGPGDSRLPAVQEWCAQEPVHLDSPAIKHQ 80
 DB 169 LPPDARDGNLPAAYNRLESP-----DPAVER 196

RESULT 45
 FTSY_MYCPN STANDARD: PRT: 348 AA.
 ID AC P75362;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CELL DIVISION PROTEIN FTSY HOMOLOG.
 GN FTSY OR MPN425 OR MP416.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscus;
 OC Mycoplasmales; Mycoplasma.
 OX NCBI_TaxID: 2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:ATCC 29342 / M129;
 RX MEDLINE:97105885; PubMed:8948633;
 RA Himmelfeich R., Hilbert H., Plegens H., Plrkl E., Ll B.-C.,
 RA Hiertmann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae.";
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- FUNCTION: FUNCTIONAL HOMOLOG OF SRP RECEPTOR. PROBABLY INVOLVED IN
 CC THE RECEPTION AND INSERTION OF A SUBSET OF PROTEINS AT THE
 CC MEMBRANE (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000040; AAB96064.1; -
 DR HSSP: P10121; 1FTS.
 DR InterPro: IPR000897; SRP54.
 DR Pfam: PF00448; SRP54; 1.
 DR ProDom: PD000819; SRP54; 1.
 DR ProSITE: PS00300; SRP54; 1.
 KW Signal recognition particle; GTP-binding; RNA-binding; Membrane;

KW Cell division; Complete proteome.
 FT NP_BIND 143 150 GTP (BY SIMILARITY).
 FT NP_BIND 225 229 GTP (BY SIMILARITY).
 FT NP_BIND 289 292 GTP (BY SIMILARITY).
 SQ SEQUENCE 348 AA: 38775 MW: 8A14C4DB9E1EAE29 CRC64;

Query Match 9.2%: Score 61; DB 1; Length 348;
 Best Local Similarity 25.0%; Pred. No. 33;
 Matches 25; Conservative 23; Mismatches 36; Indels 16; Gaps 6;

OY 28 OXSLMASEHKLKLGCDADVPGPDSRLPAVQEWCAQEPVHLDSPAIKHQELLTGDT 87
 DB 182 QLAEMAQ-----RIG-CDIVLPNKEET--PAVIFRGVQGIQNE-----YDFVLC-DT 226
 OY 88 QGRYRCRSGSLTGWXQLSKLE-LTGPKVLACSLADGAS 126
 DB 227 SGRIQNKTNLMELKRIYQIVQKVSASAKPOETLLVLDGTT 266

Search completed: January 7, 2002, 16:50:33
 Job time: 275 sec

REFERENCE 1 (bases 1 to 457)
 AUTHORS Hiller, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B., Chissee, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Riklin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevas, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (3), 807-828 (1996)

TITLE JOURNAL MEDLINE
 97044478
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: eslevat@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 1951 Std Error: 0.00
 Seq primer: mob, REGA+ET
 High quality sequence stop: 267.
 Location/Qualifiers
 1. 457
 /organism="Homo sapiens"
 /db_xref="GDB:1252138"
 /db_xref="taxon:9606"
 /clone="IMAGE:308725"
 /clone_lib="Soares fetal_lung_NbH119W"
 /lab_host="DH10B (ampicillin resistant)"
 /dev_stage="19 weeks"
 /note="Organ: lung; Vector: pTZ19 (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-GTGTACCATCTGTAAGTGGAGCGCCGCAATTTTCTTTTCTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTZ19 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NbH119W."

BASE COUNT 90 a 136 c 129 g 91 t 11 others
 ORIGIN

alignment_scores:
 Quality: 87.00 Length: 16
 Ratio: 5.438 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-471-276-831_COPY_1_16 x W25099 ..
 Align seg 1/1 to: W25099 from: 1 to: 457

seq_name: gb_est1:AL531425

seq_documentation_block:
 LOCUS AL531425 612 bp mRNA EST 13-FEB-2001
 DEFINITION AL531425 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSODM002YF08 5
 prime, mRNA sequence.
 ACCESSION AL531425
 VERSION AL531425.1 GI:12794918
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 612)
 AUTHORS Li, M.B., Gruber, C., Jessup, J. and Polayes, D.
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1. 612
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSODM002YF08"
 /clone_lib="LTI_NFL001_NBC4"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 112 a 187 c 189 g 121 t 3 others
 ORIGIN

alignment_scores:
 Quality: 87.00 Length: 16
 Ratio: 5.438 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-471-276-831_COPY_1_16 x AL531425 ..
 Align seg 1/1 to: AL531425 from: 1 to: 612

seq_name: gb_est2:BG389779

seq_documentation_block:
 LOCUS BG389779 712 bp mRNA EST 12-MAR-2001
 DEFINITION 602415155F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4523701 5',
 mRNA sequence.
 ACCESSION BG389779
 VERSION BG389779.1 GI:13283215
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 712)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov

Plate: LLM10426 row: m column: 14
High quality sequence stop: 708.

FEATURES

source

Location/Qualifiers

1..712

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4523701"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

BASE COUNT 124 a 240 c 194 g 154 t
ORIGIN

alignment_scores:

Quality: 68.50 Length: 17
Ratio: 4.281 Gaps: 1
Percent Similarity: 94.118 Percent Identity: 82.353

alignment_block:

US-09-471-276-831_COPY_1_16 x BG389779/rev ..

Align seq 1/1 to reverse of: BG389779 from: 1 to: 712

1 MetSerMetLeuValAlpheLeuLeuTrpGly...ValThrTrpG1 16
|||||
137 ATGTCCATGCTGCTGCTTCTTCTGCTGCGGAGGCGCTCCTGGG 88
16 y 16
87 T 87

seq_name: gb_est1:AM252240

seq_documentation_block:

LOCUS AM252240 259 bp mRNA EST 17-DEC-1999
DEFINITION UI-R-BJ0-dt-d-11-0-UI-s1 UI-R-BJ0 Rattus norvegicus cDNA clone
UI-R-BJ0-dt-d-11-0-UI 3', mRNA sequence.

ACCESSION AM252240
VERSION AM252240.1 GI:6595871

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 259)

AUTHORS Bonaldi,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL MEDLINE

COMMENT 97044477

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: mssoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized AV canal at 15 dpc library cDNA library preparation:
M.B. Soares Lab clone distribution: clones will be available
through Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 1-41.

>POLY A#Simple_repeat
Seq primer: M13 Forward
POLY A=yes

FEATURES

source

Location/Qualifiers

1..259

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ0-dt-d-11-0-UI"
/clone_lib="UI-R-BJ0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0
library is a subtracted library derived from the UI-R-AA1,
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the Not I site
and the oligo-dT track. The library was constructed as
described by Bonaldi, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG_LIB=UI-R-BJ0
TAG_TISSUE=AV canal at 15 dpc
TAG_SEQ=GAMAG

BASE COUNT 79 a 48 c 48 g 84 t
ORIGIN

alignment_scores:

Quality: 60.00 Length: 13
Ratio: 5.000 Gaps: 0
Percent Similarity: 92.308 Percent Identity: 69.231

alignment_block:

US-09-471-276-831_COPY_1_16 x AM252240/rev ..

Align seq 1/1 to reverse of: AM252240 from: 1 to: 259

4 LeuValAlpheLeuLeuTrpGlyValThrTrpGly 16
|||||
99 TTATATCGTTTGTGTTGTTGGGAGTACTTGGGG 61

seq_name: gb_est1:BB048712

seq_documentation_block:

LOCUS BB048712 229 bp mRNA EST 25-JUN-2000
DEFINITION BB048712 RIKEN full-length enriched, adult male olfactory bulb Mus
musculus cDNA clone 6430596408 3', mRNA sequence.

ACCESSION BB048712
VERSION BB048712.1 GI:8455860

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 229)

AUTHORS

Kono,H., Aizawa,K., Akahira,S., Akiyama,J., Arahawa,T., Carlncl
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadoya,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomioka,N., Taya
T., Tsunoda,Y., Watanabe,A., Watanabe,S., Yamamura,T., Yamazaki,I.,
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Kono,H., et al.)
Unpublished (2000)

JOURNAL

AUTHORS Xiao, H.S., Han, Z.G., Zhang, F.X., Huang, Q.H., Lu, Y.J., Bao, L., Fu, G., Guo, C., Yan, Q., Jin, S.X., Zhu, Z.D., Xu, X.R., Li, N.G., Chen, Z. and Zhang, X.

TITLE Distinct gene expression profiles of rat dorsal root ganglion induced by peripheral nerve axotomy

JOURNAL Unpublished (2001)

COMMENT Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn

This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)

PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA-No.

FEATURES
source location/Qualifiers
1..581
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRA05603"
/clone_id="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"

BASE COUNT 147 a 175 c 115 g 140 t 4 others

ORIGIN

alignment_scores:
Quality: 55.00 Length: 11
Ratio: 5.500 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 81.818

alignment_block:
US-09-471-276-831_COPY_1_16 x BG662957/rev ..

Align seg 1/1 to reverse of: BG662957 from: 1 to: 581

6 ValPheLeuLeuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||
568 GTTTTGTGTTTGTGTCGGGTGTGCGCGG 536

seq_name: gb_gss:AZ675056

seq_documentation_block:
LOCUS AZ675056 1031 bp DNA GSS 14-DEC-2000
DEFINITION ENTHH087FB Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ675056
VERSION AZ675056.1 GI:11812202
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Eumetazoa; Excavates; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 1031)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: enta@igr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 84
High quality sequence stop: 309.

FEATURES
source location/Qualifiers
1..1031
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_id="Entamoeba histolytica Sheared DNA"
/note="Vector: pHD31, Site 1, Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 121 a 113 c 393 g 404 t

ORIGIN

alignment_scores:
Quality: 55.00 Length: 13
Ratio: 5.000 Gaps: 0
Percent Similarity: 84.615 Percent Identity: 69.231

alignment_block:
US-09-471-276-831_COPY_1_16 x AZ675056 ..

Align seg 1/1 to: AZ675056 from: 1 to: 1031

4 LeuValPheLeuLeuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||
600 TTGTCGTGTTTGTGTCGGGTGTGCGCGGT 638

seq_name: gb_est1:A1712777

seq_documentation_block:
LOCUS A1712777 136 bp mRNA EST 08-JUN-1999
DEFINITION UI-R-AFL-aax-d-07-0-UI s1 UI-R-AFL Rattus norvegicus cDNA clone
ACCESSION A1712777
VERSION A1712777.1 GI:5016577
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 136)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mssoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized AV canal at 15 dpc library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-41,
>POLY_A#Simple.repeat
Seq primer: M13 Forward
POLY_A=yes.

FEATURES

source

Location/Qualifiers
1. .136
/organism:"Rattus norvegicus"
/strain:"Sprague-Dawley"
/db_xref:"taxon:10116"
/clone:"UT-R-AFI-aax-d-07-0-UI"
/clone_lib:"UT-R-AFI"
/dev_stage:"adult"
/lab_host:"DH10B (Life Technologies)"
/note:"Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; The UT-R-AFI library is a normalized library constructed from 15 dpc rat atriocentricular (AV) canal. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Jim Lin, Department of Biology, University of Iowa.
TAG_Lib-UI-R-AFI
TAG_Tissue-AV canal at 15 dpc
TAG_SEQ-GAAGC"

BASE COUNT 48 a 30 c 10 g 48 t

ORIGIN

alignment_scores:
Quality: 54.00 Length: 10
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:

US-09-471-276-831_COPY_1_16 x A1712777/rev ..

Align seg 1/1 to reverse of: A1712777 from: 1 to: 136

7 Phleuleuleutrpcllyvalthrttpgly 16
|||||:::|||||
90 tttttgtttgtttggcagcttacttggcgc 61

seq_name: gb_est1:A1010058

seq_documentation_block:

LOCUS A1010058 252 bp mRNA EST 15-JUN-1998
DEFINITION EST204509 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
RUBRT10 3' end, mRNA sequence.
A1010058
A1010058.1 GI:3223890
EST.

ACCESSION VERSION KEYWORDS SOURCE

EST.

Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 252)
Lee,N.H., Glodok,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat

TITLE

Unpublished (1998)

JOURNAL

Comment: The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529

Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-2L.

FEATURES

source

Location/Qualifiers
1. .252
/organism:"Rattus sp."
/db_xref:"taxon:10116"
/clone:"RUBRT10"
/clone_lib:"Normalized rat lung, Bento Soares"
/note:"Organ: Lung; Vector: pRT3D-Pac; Site_1: EcoRI;
Site_2: NotI"

BASE COUNT 89 a 51 c 58 g 54 t

ORIGIN

alignment_scores:
Quality: 54.00 Length: 10
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:

US-09-471-276-831_COPY_1_16 x A1010058/rev ..

Align seg 1/1 to reverse of: A1010058 from: 1 to: 252

7 Phleuleuleutrpcllyvalthrttpgly 16
|||||:::|||||
47 tttttgtttgtttggcagcttacttggcgc 18

seq_name: gb_est1:AW253137

seq_documentation_block:

LOCUS AW253137 255 bp mRNA EST 17-DEC-1999
DEFINITION UT-R-BD0-aet-a-03-0-UI.s1 UT-R-BD0 Rattus norvegicus cDNA clone
UT-R-BD0-aet-a-03-0-UI 3', mRNA sequence.
AW253137
AW253137.1 GI:6596728
EST.

KEYWORDS SOURCE

EST.

Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 255)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)

JOURNAL

Medline
97044477

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@iucw.wisc.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized ventricle at 16.5 dpc library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-41,
>POLY_A#Simple.repeat
Seq primer: M13 Forward
POLY_A=yes.

FEATURES

source

Location/Qualifiers
1. .255
/organism:"Rattus norvegicus"
/strain:"Sprague-Dawley"
/db_xref:"taxon:10116"


```

/clone="UI-R-BJ0-aet-a-03-0-UI"
/clone_lib="UI-R-BJ0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0
library is a subtracted library derived from the UI-R-A1,
UI-R-A1, UI-R-A1, UI-R-A1, UI-R-A1, UI-R-A1, and
UI-R-A1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present at the Not I site
and the oligo-dr track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG_Lib=UI-R-BJ0
TAG_TISSUE=ventricle at 16.5 dpc
TAG_SEQ=GTTCG"

BASE COUNT      77 a      49 c      46 g      83 t
ORIGIN

alignment_scores:
  Quality:      54.00      Length:      10
  Ratio:        5.400      Gaps:      0
  Percent Similarity: 100.000      Percent Identity: 80.000

alignment_block:
US-09-471-276-831_COPY_1_16 x AW253137/rev ..
Align seg 1/1 to reverse of: AW253137 from: 1 to: 255

7 PheuleuleutRatPolyaThrttggly 16
|||||:|||||:|||||:|||||:|||||
90 TTTCTGTTGTTGGCGACTTGGCGG 61

seq_name: gb_estl:A1176065
seq_documentation_block:
LOCUS      A1176065      320 bp      mRNA      EST      20-JAN-1999
DEFINITION EST219642 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
ROVBR09.3' end, mRNA sequence.
ACCESSION  A1176065
VERSION    A1176065.1 GI:3726703
KEYWORDS
SOURCE
ORGANISM   Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 320)
AUTHORS  Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
Unpublished (1998)
COMMENT   Other_ESTs: TC52105
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
            Location/Qualifiers
                1..320
                /organism="Rattus sp."
                /db_xref="ATCC (Inhost):2031135"
                /db_xref="taxon:10118"
                /clone="ROVBR09"
                /clone_lib="Normalized rat ovary, Bento Soares"
                /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;

```

```

            Site_2: NotI"
BASE COUNT      99 a      61 c      71 g      89 t
ORIGIN

alignment_scores:
  Quality:      54.00      Length:      10
  Ratio:        5.400      Gaps:      0
  Percent Similarity: 100.000      Percent Identity: 80.000

alignment_block:
US-09-471-276-831_COPY_1_16 x A1176065/rev ..
Align seg 1/1 to reverse of: A1176065 from: 1 to: 320

7 PheuleuleutRatPolyaThrttggly 16
|||||:|||||:|||||:|||||:|||||
86 TTTTCTGTTGTTGGCGACTTGGCGG 57

seq_name: gb_estl:AA956508
seq_documentation_block:
LOCUS      AA956508      364 bp      mRNA      EST      04-JUL-1999
DEFINITION UI-R-E1-fk-g-09-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA clone
UI-R-E1-fk-g-09-0-UI.3' similar to gi128268621emb|AL021330|HS244F1
Human DNA sequence from PAC 244f1 on chromosome 6q16.1-16.3.
Contains nucleophosmin (nucleolar phosphoprotein, B23, NPML,
numatrin) like pseudogene, ESTs and an STS with a CAG repeat
polymorphism, mRNA sequence.
ACCESSION  AA956508
VERSION    AA956508.1 GI:4238982
KEYWORDS
SOURCE
ORGANISM   Norway rat.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 364)
AUTHORS  Bonaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
COMMENT   On May 7, 1998 this sequence version replaced gi:3120203.
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: mssoares@iue.wieg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dr track served to identify it as a clone from the normalized
            adult 12-day-embryo library. cDNA library preparation: M. Patina
            Bonaldo, Ph.D. Clone distribution: clones will be available through the
            Research Genetics This clone is also available through the
            I.M.A.G.E. Consortium at LNL (InfoImage.Lnl.gov). IMAGE
            ID=1779665 The following repetitive elements were found in this
            cDNA sequence: 1-32, >POLY_A$Simple_repeat
            Seq primer: M13 Forward
            POLY-A-No.
            Location/Qualifiers
                1..364
                /organism="Rattus norvegicus"
                /strain="Sprague-Dawley"
                /db_xref="taxon:10116"
                /clone="UI-R-E1-fk-g-09-0-UI"
                /clone_lib="UI-R-E1"
                /dev_stage="adult"
                /lab_host="DH10B (Life Technologies)"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-E1
                library is a subtracted library derived from the UI-R-E0

```

library. The UI-R-E0 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-E1) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-E0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-E0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-E1 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996).

alignment_scores:
Quality: 54.00 Length: 10
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:

US-09-471-276-831_COPY_1_16 x AA956508/rev ..

Align seg 1/1 to reverse of: AA956508 from: 1 to: 364

7 PhleuleuleuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||:|||||
81 TTTTCTTTGTTTGGCGACTTACTTGGCGC 52

seq_name: gb_est1:A1764098

seq_documentation_block: 366 bp mRNA 25-JUN-1999
LOCUS A1764098
DEFINITION UI-R-Y0-acy-g-11-0-UI-81 UI-R-Y0 Rattus norvegicus cDNA clone
UI-R-Y0-acy-g-11-0-UI 3', mRNA sequence.
ACCESSION A1764098
VERSION A1764098.1 GI:5210033
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 366)
Bonaldi,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonaldi poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized Eye library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA+yes.

FEATURES
source

Location/Qualifiers

1..366
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-acy-g-11-0-UI"
/clone_lib="UI-R-Y0"
/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-Y0 library is a subtracted library derived from an individually-tagged normalized whole-eye (minus the lens) library. The driver for the subtraction consisted of a pool of all previous libraries (UI-R-Y0, UI-R-Y1, UI-R-E0, UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996)

BASE COUNT 117 a 62 c 76 g 111 t
ORIGIN
TAG_LIB=UI-R-Y0
TAG_TISSUE=EYE
TAG_SEQ_CAT=TC

alignment_scores:
Quality: 54.00 Length: 10
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:

US-09-471-276-831_COPY_1_16 x A1764098/rev ..

Align seg 1/1 to reverse of: A1764098 from: 1 to: 366

7 PhleuleuleuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||:|||||
102 TTTTCTTTGTTTGGCGACTTACTTGGCGG 73

seq_name: gb_est1:A1408881

seq_documentation_block: 375 bp mRNA 09-FEB-1999
LOCUS A1408881
DEFINITION EST237172 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
ROVER35 3' end, mRNA sequence.
ACCESSION A1408881
VERSION A1408881.1 GI:4252385
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 375)
Lee,N.H., Glodde,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529
Fax: (301)-838-0208

Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
source Location/Qualifiers

1..375
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="R0VER35"
/note="Organ: ovary; Vector: pT73pac; Site_1: EcoRI;
Site_2: NotI"

BASE COUNT 123 a 61 c 95 g 96 t
ORIGIN

alignment_scores:
Quality: 54.00 Length: 10
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:
US-09-471-276-831_COPY_1_16 x A1408881/rev ..

Align seg 1/1 to reverse of: A1408881 from: 1 to: 375

7 Phleuleuleutrpcllyvaltrtpgly 16
|||||:|||||:|||||:|||||:|||||
74 TTTTGTCTTGTGGGAGTACTCGCGG 45

seq_name: gb_est1:A1101493

seq_documentation_block:
LOCUS A1101493 379 bp mRNA EST 31-JAN-1999

DEFINITION EST210782 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
RBRBP87 3' end, mRNA sequence.

ACCESSION A1101493
VERSION A1101493.1 GI:3706384

KEYWORDS
SOURCE
ORGANISM

Rattus sp.
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 379)

AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.

TITLE Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index

JOURNAL Unpublished (1998)
COMMENT Other_ESTs: TC52105

The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
source Location/Qualifiers

1..379
/organism="Rattus sp."
/db_xref="ATCC (Inhost):2024108"
/db_xref="taxon:10118"
/clone="RBRBP87"
/clone_lib="Normalized rat brain, Bento Soares"
/note="Organ: brain; Vector: pT73pac; Site_1: EcoRI;
Site_2: NotI"

BASE COUNT 122 a 61 c 93 g 103 t
ORIGIN

alignment_scores:
Quality: 54.00 Length: 10
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:
US-09-471-276-831_COPY_1_16 x A1101493/rev ..

Align seg 1/1 to reverse of: A1101493 from: 1 to: 379

7 Phleuleuleutrpcllyvaltrtpgly 16
|||||:|||||:|||||:|||||:|||||
74 TTTTGTCTTGTGGGAGTACTCGCGG 45

seq_name: gb_est1:A1704632

seq_documentation_block:

LOCUS A1704632 388 bp mRNA EST 03-JUN-1999
DEFINITION UI-R-AB1-YW-d-06-0-UI.s1 UI-R-AB1 Rattus norvegicus cDNA clone
UI-R-AB1-YW-d-06-0-UI 3', mRNA sequence.

ACCESSION A1704632
VERSION A1704632.1 GI:4992532

KEYWORDS
SOURCE
ORGANISM

Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 388)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE
COMMENT

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized ventricle at 16.5 dpc library cDNA library preparation:

M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 1-41.
>POLY_A#Simple-repeat
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source

Location/Qualifiers
1..388
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-AB1-YW-d-06-0-UI"
/clone_lib="UI-R-AB1"
/dev_stage="adult"
/lab_host="PH10B (Life Technologies)"
/note="Vector: pT7T3D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AB1
library is a normalized library constructed from 16.5 dpc
rat ventricle. The tag is a string of 5 nucleotides
present between the Not I site and the oligo-dT track.
The library was constructed as described by Bonaldo,
Lennon and Soares, Genome Research 6: 791-806, 1996.
Tissue provided by Jim Lin, Department of Biology,
University of Iowa.
TAG_LIB=UI-R-AB1
TAG_TISSUE=ventricle at 16.5 dpc

BASE COUNT 122 a 62 c 89 g 115 t
ORIGIN
TAG_SEQ:GTTCC"

Alignment_scores:
Quality: 54.00 Length: 10
Ratio: 5.400 Caps: 0
Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:
US-09-471-276-831_COPY_1_16 x A1704632/rev ..

Align seg 1/1 to reverse of: A1704632 from: 1 to: 388

7 PhleuleuleuTrpGlyValThrTrpGly 16
|||||:.....:|||||:|||||:|||||
90 TTTTGTGTTGTTGGGAGTACTTGCGG 61

seq_name: gb_est1:AW251200

seq_documentation_block:
LOCUS AW251200 390 bp mRNA EST 17-DEC-1999
DEFINITION UI-R-BJ0-adj-d-02-0-UI.s1 UI-R-BJ0 Rattus norvegicus cDNA clone
ACCESSION AW251200
VERSION AW251200.1 GI:6594791
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 390)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704447
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized AV canal at 15 dpc library cDNA library Preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 1-41,
>POLY_A#Simple.repeat
Seq primer: M13 Forward
POLYA=yes.

FEATURES
Source
Location/Qualifiers
1..390
/organism:"Rattus norvegicus"
/strain:"Sprague-Dawley"
/db_xref:"taxon:10116"
/clone:"UI-R-BJ0-adj-d-02-0-UI"
/dev_stage:"adult"
/clone_1lb:"UI-R-BJ0"
/lab_host:"DH10B (Life Technologies)"
/note:"Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0
library is a subtracted library derived from the UI-R-AA1,
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal

at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the NotI site
and the oligo-dT track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG_LIB UI-R-BJ0
TAG_TISSUE:AV canal at 15 dpc
TAG_SEQ GAAGC"

BASE COUNT 123 a 62 c 92 g 113 t
ORIGIN

Alignment_scores:
Quality: 54.00 Length: 10
Ratio: 5.400 Caps: 0
Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:
US-09-471-276-831_COPY_1_16 x AW251200/rev ..

Align seg 1/1 to reverse of: AW251200 from: 1 to: 390

7 PhleuleuleuTrpGlyValThrTrpGly 16
|||||:.....:|||||:|||||:|||||
90 TTTTGTGTTGTTGGGAGTACTTGCGG 61

seq_name: gb_est1:BE109230

seq_documentation_block:
LOCUS BE109230 393 bp mRNA EST 13-JUN-2000
DEFINITION UI-R-BS1-azd-d-09-0-UI.s1 UI-R-BS1 Rattus norvegicus cDNA clone
ACCESSION BE109230
VERSION BE109230.1 GI:8501335
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 393)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704447
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized embryo at 13 dpc library cDNA library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.

FEATURES
Source
Location/Qualifiers
1..393
/organism:"Rattus norvegicus"
/strain:"Sprague-Dawley"
/db_xref:"taxon:10116"
/clone:"UI-R-BS1-azd-d-09-0-UI"
/clone_1lb:"UI-R-BS1"
/dev_stage:"embryonic 13 dpc"
/lab_host:"DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-BS1 library is derived from 13 dpc whole embryo tissue. For a detailed description of the library from which this clone was derived, please visit our web site at rblast.eng.uiowa.edu.

TAG-LIB-UI-R-BS1
TAG-TISSUE-embryo at 13 dpc
TAG-SEQ-AATCC"

BASE COUNT

123 a 64 c 89 g 117 t

alignment_scores:
Quality: 54.00 Length: 10
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:

US-09-471-276-831_COPY_1_16 x BE109230/rev ..

Align seg 1/1 to reverse of: BE109230 from: 1 to: 393

7 PheleuleuleuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||:|||||:
102 TTTTCTTTGTTGGGAGTACTTGGGGG 73

seq_name: gb_gss:AQ112808

seq_documentation_block:

LOCUS AQ112808 438 bp DNA GSS 29-AUG-1998
DEFINITION CIT-HSP-2382X1.TF CIT-HSP Homo sapiens genomic clone 2382X1, DNA sequence.

ACCESSION AQ112808
VERSION AQ112808.1 GI:3488929
KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 438)
Adams,M.D., Rounsley,S.D., Zhao,S., Baas,S., Linher,K., Golden,K.,

Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and

Venter,J.C.

Map Building

Unpublished (1998)

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: M13-21

Class: BAC ends.

FEATURES

source

1. .438
Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_id="2382X1"

/clone_id="CIT-HSP"

/sex="Male"

/cell_type="Sperm"

/note="Vector: pbelOBAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 130 a 100 c 84 g 124 t

alignment_scores:

Quality: 54.00 Length: 11
Ratio: 5.400 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 72.727

alignment_block:

US-09-471-276-831_COPY_1_16 x AQ112808/rev ..

Align seg 1/1 to reverse of: AQ112808 from: 1 to: 438

6 ValPheleuleuleuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||:|||||:
144 GTTACCTTTACTTGGGATTTGGGGG 112

seq_name: gb_estl:AW251345

seq_documentation_block:

LOCUS AW251345 470 bp mRNA EST 17-DEC-1999
DEFINITION UI-R-BJ0-adh-a-11-0-UI.s1 UI-R-BJ0 Rattus norvegicus cDNA clone

UI-R-BJ0-adh-a-11-0-UI 3', mRNA sequence.

ACCESSION AW251345

VERSION AW251345.1 GI:6594936

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 470)

Bonaldi,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL MEDLINE

97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the

normalized AV canal at 15 dpc library cDNA library preparation:

M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-41,
>POLY_A\$imple_repeat
Seq primer: M13 Forward
POLY-A\$es.

FEATURES

source

1. .470
Location/Qualifiers

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone_id="UI-R-BJ0-adh-a-11-0-UI"

/clone_id="UI-R-BJ0"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0

library is a subcloned library derived from the UI-R-A41,

UI-R-A81, UI-R-A41, UI-R-A41, UI-R-A41, UI-R-A41, and

UI-R-A41 libraries. These libraries represent tissues from

rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal

at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV

canal at 15 dpc, and ventricle at 13 dpc. The tag is a

string of 5-6 nucleotides present between the Not I site

and the oligo-dT track. The library was constructed as

described by Bonaldi, Lennon and Soares, Genome Research

```
6: 791-806, 1996.
TAG_LIB: UI-R-BJ0
TAG_TISSUE: AV canal at 15 dpc
TAG_SEQ: GAAGC*
BASE COUNT      156 a      76 c      104 g      134 t
ORIGIN

alignment_scores:
  Quality: 54.00      Length: 10
  Ratio: 5.400      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 80.000

alignment_block:
US-09-471-276-831_COPY_1_16 x AW251345/rev ..

Align seg 1/1 to reverse of: AW251345 from: 1 to: 470

7 PhleuleuleuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||
90 TTTTGTGTTGTTGGGAGCTTACTTGCGGG 61

seq_name: gb_est1:AW252673

seq_documentation_block:
LOCUS      AW252673      475 bp      mRNA      EST      17-DEC-1999
DEFINITION UI-R-BJ0-aeb-c-05-0-UI.s1 UI-R-BJ0 Rattus norvegicus cDNA clone
VERSION     AW252673      GI:5596264
KEYWORDS    EST.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 475)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
COMMENT    Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@iuc.wiieg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NctI site
            and the oligo-dT track served to identify it as a clone from the
            normalized ventricle at 16.5 dpc library cDNA library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 1-41,
            >POLY_A*Simple_repeat
            Seq primer: M13 Forward
            POLY-A-Ycs.

FEATURES
  source
  1..475
    /organism="Rattus norvegicus"
    /strain="Sprague-Dawley"
    /db_xref="taxon:10116"
    /clone="UI-R-BJ0-aeb-c-05-0-UI"
    /clone_lib="UI-R-BJ0"
    /dev_stage="adult"
    /lab_host="DH10B (Life Technologies)"
    /note="Vector: pT73D-Pac (Pharmacia) with a modified
    polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0
    library is a subtracted library derived from the UI-R-A1,
```

```
UI-R-A1, UI-R-Ac1, UI-R-Ad1, UI-R-AE1, UI-R-AF1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the Nct I site
and the oligo-dT track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG_LIB: UI-R-BJ0
TAG_TISSUE: ventricle at 16.5 dpc
TAG_SEQ: GTTCG*
BASE COUNT      159 a      76 c      106 g      134 t
ORIGIN

alignment_scores:
  Quality: 54.00      Length: 10
  Ratio: 5.400      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 80.000

alignment_block:
US-09-471-276-831_COPY_1_16 x AW252673/rev ..

Align seg 1/1 to reverse of: AW252673 from: 1 to: 475

7 PhleuleuleuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||
90 TTTTGTGTTGTTGGGAGCTTACTTGCGGG 61

seq_name: gb_est1:A1176739

seq_documentation_block:
LOCUS      A1176739      479 bp      mRNA      EST      20-JAN-1999
DEFINITION EST220332 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
VERSION     A1176739      GI:3727377
KEYWORDS    EST.
SOURCE      Rattus sp.
ORGANISM    Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 479)
AUTHORS    Lee,N.H., Glodex,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE      Rat Genome Project: Generation of a Rat EST (RESt) Catalog & Rat
            Gene Index
JOURNAL    Unpublished (1998)
COMMENT    Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.

FEATURES
  source
  1..479
    /organism="Rattus sp."
    /db_xref="ATCC (Inhost):2031503"
    /db_xref="taxon:10118"
    /clone="ROVBM23"
    /clone_lib="Normalized rat ovary, Bento Soares"
    /note="Organ: ovary; Vector: pT73D-Pac; Site_1: EcoRI;
    Site_2: NctI"
BASE COUNT      159 a      77 c      107 g      136 t
ORIGIN

alignment_scores:
  Quality: 54.00      Length: 10
  Ratio: 5.400      Gaps: 0
```

Percent Similarity: 100.000 Percent Identity: 80.000

alignment_block:

US-09-471-276-831_COPY_1_16 x A1176739/rev ..

Align seg 1/1 to reverse of: A1176739 from: 1 to: 479

7 PhleuleuleuTripglyValThrTpgly 16
|||||:.....:|||||:|||||:|||||
86 TTTTGTGTTGGGAGTACTGCGG 57

seq_name: gb_gss:A2776969

seq_documentation_block:

LOCUS A2776969 550 bp DNA GSS 16-FEB-2001
DEFINITION 2M0011K05F Mouse 10kb plasmid U0GC1M library Mus musculus genomic
clone U0GC2M0011K05 F, DNA sequence.

ACCESSION A2776969

VERSION A2776969.1 GI:12905099

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Maimoud,M., Meenen,E., Pedersen,T., Reilly,
'M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A.,
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)

TITLE

JOURNAL

COMMENT

Contact: Robert B. Weiss
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0011 row: K column: 05
Seq primer: CTTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 550.
Location/Qualifiers

FEATURES
SOURCE

1..550
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U0GC2M0011K05"
/clone_lib="Mouse 10kb plasmid U0GC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g114732114gblAF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 143 a 108 c 139 g 159 t 1 others
ORIGIN

alignment_scores:

Quality: 54.00 Length: 12
Ratio: 5.400 Caps: 0
Percent Similarity: 83.333 Percent Identity: 75.000

alignment_block:

US-09-471-276-831_COPY_1_16 x A2776969 ..

Align seg 1/1 to: A2776969 from: 1 to: 550

5 ValValPhleuleuleuTripglyValThrTpgly 16
|||||:|||||:|||||:|||||
159 GTTGCTCTCCCTGTGATTGGCGCTGCGAGTGGGC 194

seq_name: gb_est2:B1298800

seq_documentation_block:

LOCUS B1298800 433 bp mRNA EST 20-JUL-2001
DEFINITION UI-R-CV2-chn-g-10-0-UI.s1 UI-R-CV2 Rattus norvegicus cDNA clone
UI-R-CV2-chn-g-10-0-UI 3', mRNA sequence.

ACCESSION B1298800

VERSION B1298800.1 GI:14975080

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

1 (bases 1 to 433)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msquares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized rat eye library cDNA library preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com) The following repetitive elements were
found in this cDNA sequence: 1-42, >AT-rich#low_complexity
Seq primer: M13 Forward
POLYA=yes.
Location/Qualifiers

FEATURES
SOURCE

1..433
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CV2-chn-g-10-0-UI"
/clone_lib="UI-R-CV2"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; CV2 is a
subtracted library derived from a normalized eye library
(nRE22), constructed according to the procedure described
by Bonaldo, Lennon & Soares (Genome Research Genome 6:
791-806, 1996). For construction of the CV2 library,
plasmid DNA from the normalized rat eye library was
electroporated into competent bacteria for production of

single-stranded circular DNA. This was then used as a tracer in a subtractive hybridization with a driver comprising: a) a set of about 1,000 arrayed clones from the non-normalized eye library CVO (PCR amplified inserts from a plasmid DNA template preparation representing plates R-CV0-BR8 through R-CV0-BR8). This represented 20% of the final driver population. b) A set of about 4,000 arrayed clones from the normalized eye library CV1 (PCR amplified inserts from a plasmid DNA template preparation representing plates R-CV1-BR8 through R-CV1-BR8 through R-CV1-BR8, and R-CV1-BV0 through R-CV1-BV0). This represented 80% of the final driver population.

TAG_L1B: UI-R-CV2
TAG_TISSUE: rat eye
TAG_SEQ: CAGCCC*

BASE COUNT 161 a 63 c 60 g 149 t
ORIGIN

alignment_scores:
Quality: 53.00 Length: 13
Ratio: 4.417 Gaps: 0
Percent Similarity: 92.308 Percent Identity: 61.538

alignment_block:
US-09-471-276-831_COPY_1_16 x B1298800/rev ..

Align seg 1/1 to reverse of: B1298800 from: 1 to: 433

3 MetLeuValValPheLeuLeuTrpGlyValThrTrp 15
300 TTAGTTGTGTTTATTTCCTTCGTAATGACATTCG 262

seq_name: gb_estc2:BF339365

seq_documentation_block:
LOCUS BF339365 741 bp mRNA EST 22-NOV-2000
DEFINITION 602038716P1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4186610
5' mRNA sequence.
ACCESSION BF339365
VERSION BF339365.1 GI:11285832
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 741)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@b-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1AM9507 row: p column: 03
High quality sequence stop: 646.
Location/Qualifiers
1..741
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/clone:"IMAGE:4186610"
/clone_lib:"NCI_CGAP_Brn64"
/tissue_type:"glioblastoma with EGFR amplification"
/lab_host:"DH10B (TI phage-resistant)"
/note:"Organ: brain; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.57 Kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

FEATURES
source

BASE COUNT 162 a 204 c 204 g 171 t
ORIGIN

alignment_scores:
Quality: 53.00 Length: 11
Ratio: 5.889 Gaps: 0
Percent Similarity: 81.818 Percent Identity: 72.727

alignment_block:
US-09-471-276-831_COPY_1_16 x BF339365 ..

Align seg 1/1 to: BF339365 from: 1 to: 741

6 ValPheLeuLeuLeuTrpGlyValThrTrpGly 16
662 ATTTCTGCTTATGGGGCTCGACGTGGGC 694

seq_name: gb_estc2:BF673019

seq_documentation_block:
LOCUS BF673019 872 bp mRNA EST 21-DEC-2000
DEFINITION 602152912P1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293939 5',
mRNA sequence.
ACCESSION BF673019
VERSION BF673019.1 GI:11946914
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 872)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@b-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1CM1142 row: h column: 04
High quality sequence stop: 614.
Location/Qualifiers
1..872
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/clone:"IMAGE:4293939"
/clone_lib:"NIH_MGC_81"
/lab_host:"DH10B (TI phage-resistant)"
/note:"Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgccgcgcgcgc); Site_2: SfiI (ggcgcatatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

FEATURES
source

BASE COUNT 285 a 212 c 204 g 171 t
ORIGIN

alignment_scores:
Quality: 53.00 Length: 13
Ratio: 4.417 Gaps: 0
Percent Similarity: 92.308 Percent Identity: 61.538

alignment_block:
US-09-471-276-831_COPY_1_16 x BF673019/rev ..
Align seg 1/1 to reverse of: BF673019 from: 1 to: 872

4 LeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||:|||||
734 CTCGTGTTCTTCTTCCTTATGGGCGCTGTGGCGA 696

seq_name: gb_gss:CNS02MAM

seq_documentation_block:
LOCUS CNS02MAM 931 bp DNA GSS 14-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
149E11 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL203873.1 GI:7862692
VERSION AL203873.1
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 931)
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
TITLE 2 (bases 1 to 931)
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
REFERENCE Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
TITLE 3 (bases 1 to 931)
JOURNAL Genoscope.
REFERENCE Direct Submission
AUTHORS Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source 1..931
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="149E11"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG149AC06LPI-end : T7"
BASE COUNT 222 a 212 c 242 g 251 t 4 others
ORIGIN

alignment_scores:
Quality: 53.00 Length: 16
Ratio: 4.077 Gaps: 0
Percent Similarity: 81.250 Percent Identity: 50.000

alignment_block:
US-09-471-276-831_COPY_1_16 x CNS02MAM ..
Align seg 1/1 to: CNS02MAM from: 1 to: 931

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
:|||||:|||||:|||||:|||||:|||||
619 CTGAATGTTCTTCTTCCTTCTGCTGTGGGCGCTTGTGGGGG 666

seq_name: gb_gss:CNS043FU

seq_documentation_block:
LOCUS CNS043FU 964 bp DNA GSS 18-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
079H19 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL272739.1 GI:7994977
VERSION AL272739.1
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 964)
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
TITLE 2 (bases 1 to 964)
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
REFERENCE Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
TITLE 3 (bases 1 to 964)
JOURNAL Genoscope.
REFERENCE Direct Submission
AUTHORS Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source 1..964
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="079H19"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG079C010LPI-end : T7"
BASE COUNT 224 a 209 c 259 g 267 t 5 others
ORIGIN

alignment_scores:
Quality: 53.00 Length: 16
Ratio: 4.077 Gaps: 0
Percent Similarity: 81.250 Percent Identity: 50.000

alignment_block:
US-09-471-276-831_COPY_1_16 x CNS043FU ..
Align seg 1/1 to: CNS043FU from: 1 to: 964

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
:|||||:|||||:|||||:|||||:|||||
726 CTGAATGTTCTTCTTCCTTCTGCTGTGGGCGCTTGTGGGGG 773

seq_name: gb_est1:BE108872

seq_documentation_block:
LOCUS BE108872 152 bp mRNA EST 13-JUN-2000
DEFINITION UI-R-BS1-ayz-d-04-0-UI s1 UI-R-BS1 Rattus norvegicus cDNA clone
UI-R-BS1-ayz-d-04-0-UI 3', mRNA sequence.
ACCESSION BE108872
VERSION BE108872.1 GI:8509977
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

```

REFERENCE          1 (bases 1 to 152)
AUTHORS            Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE              Normalization and subtraction: two approaches to facilitate gene
JOURNAL            Genome Res. 6 (9), 791-806 (1996)
MEDLINE            97044477
COMMENT            Contact: Soares, MB
                   Program for Rat Gene Discovery and Mapping
                   University of Iowa
                   451 Eckstein Medical Research Building Iowa City, IA 52242, USA
                   Tel: 319 335 8250
                   Fax: 319 335 9565
                   Email: msoares@blue.weeg.uiowa.edu
                   The sequence contained an oligo-dT track that was present in the
                   oligonucleotide that was used to prime the synthesis of first
                   strand cDNA and therefore this may represent a bonafide poly A
                   tail. cDNA library Preparation: M.B. Soares lab clone distribution
                   clones will be available through Research Genetics (www.resgen.com)
                   Seq primer: M13 forward
                   POLYA+yes.

FEATURES
   source          Location/Qualifiers
                   1..152
                   /organism="Rattus norvegicus"
                   /strain="Sprague-Dawley"
                   /db_xref="taxon:10116"
                   /clone="UI-R-B51-ayz-d-04-0-0-U"
                   /clone_1lb="UI-R-B51"
                   /dev_stage="embryonic 13 dpc"
                   /lab_host="DH10B (Life Technologies)"
                   /note="Vector: pRT3P-Pac (Pharmacia) with a modified
                   polylinker. Site_1: Not 1; Site_2: Eco RI; The UI-R-B51
                   library is derived from 13 dpc whole embryo tissue. For a
                   detailed description of the library from which this clone
                   was derived, please visit our web site at
                   ratest.cng.uiowa.edu.
                   TAG_SEQ=None found"

BASE COUNT          31 a      23 c      49 g      49 t

ORIGIN
TAG_SEQ=None found"

alignment_scores:
   Quality:      52.00      Length:      10
   Ratio:        5.778      Gaps:      0
   Percent Similarity: 90.000      Percent Identity: 80.000

alignment_block:
US-09-471-276-831_COPY_1_16 x BE108872      ..

Align seg 1/1  to: BE108872  from: 1  to: 152

      7 PheLeuLeuLeuTrpClyValTrpTrpGly 16
      |||:|||||
      30 TTCATTCTGCTGCGGGCGCTGCGTGGCGCA 59

seq_name: gb_est1:AV272375

seq_documentation_block:
LOCUS              AV272375          232 bp      mRNA      EST      05-NOV-1999
DEFINITION         AV272375 RIKEN full-length enriched, adult male testis (DH10B) Mus
                    musculus cDNA clone 4531424020 3' similar to AB023219 Homo sapiens
ACCESSION          AV272375
VERSION            AV272375.1  GI:6260412
KEYWORDS            EST.
SOURCE             house mouse.
ORGANISM            Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE          1 (bases 1 to 232)
AUTHORS            Kono,H., Alizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,

```

```

TITLE
JOURNAL
COMMENT

Pukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Horii,F., Kai  
Ishii,Y., Isikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai  
.C., Kawai,J., Kikuchi,N., Koijima,Y., Koya,S., Kusakabe,M.,  
Matsuyama,T., Mikai,R., Mizuno,Y., Nakamura,M., Oka,H., Okazaki,Y.,  
Owa,C., Ozawa,Y., Saito,H., Sanjo,M., Satoh,K., Shibata,K., Shibata  
.Y., Shigemoto,Y., Shiraki,T., Sugawara,Y., Sugahara,Y., Suzuki,H.,  
Suzuki,H., Takahashi,F., Tatenou,M., Tomihana,N., Tsunoda,Y.,  
Watabiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokote,T.,  
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al. 1998)  
Unpublished (1998)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel.: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Sasaki,N., Izawa,M., Watabiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,  
Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki  
.Y.  
Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Itoh.M., Kitsumai.T., Akiyama.J., Shibata.K., Izawa.M., Kawai.J.,  
Tomaru.Y., Carninci.P., Shibata.Y., Ozawa.Y., Muramatsu.M., Okazaki  
.Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for  
further details.
```

```
FEATURES
SOURCE

1..232
Location/Organism
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="_4931424O20"
/clone_lib."_RIKEN full-length enriched, adult male testis (DHIOB)"
/sex_"male"
/tissue_type_"testis"
/dev_stage_"adult"
/lab_host_"DHIOB"
/note_"Site_1: SalI; Site_2: BamHI: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGACAGAGAAGATCCAGTCAAGCTCTTTTTTTTTTTTTTTVNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGACAGATTTCGAGTAATAAATTAATGCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamH sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
```

```
BASE COUNT      47 a          46 c          46 g          93 t
ORIGIN
```

```
alignment_scores:
Quality:       52.00           Length:        11
Ratio:         5.200             Caps:         0
Percent Similarity: 90.909     Percent Identity: 72.727
```

```
alignment_block:
```


BASE COUNT 84 a 97 c 128 g 114 t
ORIGIN

alignment_scores:
Quality: 52.00 Length: 14
Ratio: 4.727 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 57.143

alignment_block:
US-09-471-276-831_COPY_1_16 x B1053983 ..

Align seg 1/1 to: B1053983 from: 1 to: 423

3 MetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
||||| ||| :|||||:|||||:|||||
210 ATGCTTGAGTGTGTATGCTGTGGTGGACTGCGACTGCGGA 251

seq_name: gb_gss:A0975452

seq_documentation_block:
LOCUS A0975452 450 bp DNA GSS 28-JAN-2000
DEFINITION RPCI-23-333110.TV RPCI-23 Mus musculus genomic clone RPCI-23-333110
ACCESSION A0975452
VERSION A0975452.1 GI:6805909
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 450)
, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other GSSs: RPCI-23-333110.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@edj.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 333 row: I column: 10
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..450
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-333110"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1;
EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 87 a 102 c 127 g 132 t 2 others

ORIGIN

alignment_scores:
Quality: 52.00 Length: 14
Ratio: 4.000 Gaps: 0
Percent Similarity: 92.857 Percent Identity: 57.143

alignment_block:
US-09-471-276-831_COPY_1_16 x A0975452 ..

Align seg 1/1 to: A0975452 from: 1 to: 450

2 SerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrp 15
:::|||||:|||||:|||||:|||||:|||||
4 CCTATGCTGTGCTTATTCGTGTGCTTTGGCGTCCACCTCG 45

seq_name: gb_est2:B1054283

seq_documentation_block:
LOCUS B1054283 471 bp mRNA EST 15-JUN-2001
DEFINITION PM3-GN0375-010201-004-B12 GN0375 Homo sapiens CDNA, mRNA sequence.
ACCESSION B1054283
VERSION B1054283.1 GI:14461813
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 471)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsumura, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
200202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rue Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?PL-PM3&t2-PM3-GN0375-
010201-004-b12&t3-2001-02-01&t4-1)
Seq primer: puc 18 forward
High quality sequence stop: 471.

FEATURES
source Location/Qualifiers
1..471
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0375"
/dev_stage="Adult"
/note="Organ: placenta; normal; Vector: puc18; Site: 1; SmaI
; Site: 2; SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the puc 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."

BASE COUNT 105 a 107 c 140 g 119 t

ORIGIN

alignment_scores:
Quality: 52.00 Length: 14
Ratio: 4.727 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 57.143

Alignment_block:

US-09-471-276-831_COPY_1_16 x BI054283 ..

Align seg 1/1 to: BI054283 from: 1 to: 471

3 MetLeuValAlpheleuleuLeuTrpGlyValThrTrpGly 16
||||| ||| :|||||
211 ATCGCTGACGTGCTGATGCTGCGACGTGCGGCA 252

seq_name: gb_gss:A0977668

seq_documentation_block:

LOCUS A0977668 538 bp DNA GSS 29-JAN-2000
DEFINITION RPCI-23-336K8-TV RPCI-23 Mus musculus genomic clone RPCI-23-336K8,
DNA sequence.ACCESSION A0977668
VERSION A0977668.1 GI:6809969

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 538)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akiret,
'B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-336K8-TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Reses ch Genetics (http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 336 row: K column: 8
Seq primer: T7
Class: BAC ends.

FEATURES
Location/Qualifiers

1..538
/organism:"Mus musculus"
/strain:"C57BL/6J"
/db_xref:"taxon:10090"
/clone:"RPCI-23-336K8"
/clone_lib:"RPCI-23"
/sex:"Female"
/lab_host:"DH10B"
/note:"Organ: Kidney/Brain; Vector: pBACe3.6; Site:1:
EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI MethyIase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 109 a 119 c 157 g 151 t 2 others
ORIGIN

Alignment_scores:

Quality: 52.00 Length: 14
Ratio: 4.000 Gaps: 0
Percent Similarity: 92.857 Percent Identity: 57.143

Alignment_block:

US-09-471-276-831_COPY_1_16 x A0977668 ..

Align seg 1/1 to: A0977668 from: 1 to: 538

2 SerMetLeuValAlpheleuleuLeuTrpGlyValThrTrp 15
:::|||||:|||||:|||||:|||||:|||||
57 GCTATCGTGTGTGTATTCGTGTGTGTGTGTGCGGCTGCGGCA 98

seq_name: gb_gss:A2097580

seq_documentation_block:

LOCUS A2097580 704 bp DNA GSS 09-MAY-2000
DEFINITION RPCI-23-15C10-TV RPCI-23 Mus musculus genomic clone RPCI-23-15C10,
DNA sequence.ACCESSION A2097580
VERSION A2097580.1 GI:7750636

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 704)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akiret,
'B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-15C10-TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Reses ch Genetics (http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 15 row: C column: 10
Seq primer: SP6
Class: BAC ends.

FEATURES
Location/Qualifiers

1..704
/organism:"Mus musculus"
/strain:"C57BL/6J"
/db_xref:"taxon:10090"
/clone:"RPCI-23-15C10"
/clone_lib:"RPCI-23"
/sex:"Female"
/lab_host:"DH10B"
/note:"Organ: Kidney/Brain; Vector: pBACe3.6; Site:1:
EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI MethyIase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 133 a 164 c 204 g 203 t
ORIGIN

Alignment_scores:

Quality: 52.00 Length: 14
Ratio: 4.000 Gaps: 0
Percent Similarity: 92.857 Percent Identity: 57.143

Alignment_block:

US-09-471-276-831_COPY_1_16 x A2097580 ..
Align seg 1/1 to: A2097580 from: 1 to: 704

AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CCAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL, send email to:
info@image.llnl.gov
Seq primer: -40bp from Glibco.
Location/Qualifiers
1..333
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/clone:"IMAGE:3319792"
/clone_lib:"NCI-CCAP_Col6"
/tissue_type:"colon tumor, RER"
/lab_host:"DH10B"
/note:"Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI-CCAP Col6 was
prepared, and as circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(clonoids 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 142 a 72 c 55 g 64 t
ORIGIN

alignment_scores:
Quality: 51.50 Length: 13
Ratio: 4.682 Gaps: 1
Percent Similarity: 84.615 Percent Identity: 76.923

alignment_block:
US-09-471-276-831_COPY_1_16 x BF591140/rev ..
Align seg 1/1 to reverse of: BF591140 from: 1 to: 333
5 ValValPheLeuLeuLeuTrpGlyValThr...TrpGly 16
|||||
242 GATTCTTTTACTACTTGGGGGATTAACGTGGTGGCG 204
seq_name: gb_est2:BF390055

seq_documentation_block:
LOCUS BF390055 433 bp mRNA EST 27-NOV-2000
DEFINITION UI-R-BS2-bdr-g-04-0-UI.s1 UI-R-BS2 Rattus norvegicus cDNA clone
UI-R-BS2-bdr-g-04-0-UI 3', mRNA sequence.
ACCESSION BF390055
VERSION BF390055.1 GI:11374897
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 433)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized embryo at 13 dpc library cDNA Library Preparation: M.B.
Soares Lab Clone Distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.
Location/Qualifiers
1..433
/organism:"Rattus norvegicus"
/strain:"Sprague-Dawley"
/db_xref:"taxon:10116"
/clone:"UI-R-BS2-bdr-g-04-0-UI"
/clone_lib:"UI-R-BS2"
/dev_stage:"embryonic 13 dpc"
/lab_host:"DH10B (Life Technologies)"
/note:"Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BS2
library is a subtracted library derived from 13 dpc whole
embryo tissue. For a detailed description of the library
from which this clone was derived, please visit our web
site at ratest.eng.uiowa.edu. The subtraction has been
previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_LIB:UI-R-BS2
TAG_TISSUE:embryo at 13 dpc
TAG_SEQ:ATATCC"
BASE COUNT 163 a 85 c 75 g 106 t 4 others
ORIGIN

alignment_scores:
Quality: 51.50 Length: 13
Ratio: 4.682 Gaps: 1
Percent Similarity: 84.615 Percent Identity: 69.231

alignment_block:
US-09-471-276-831_COPY_1_16 x BF390055/rev ..
Align seg 1/1 to reverse of: BF390055 from: 1 to: 433
5 ValValPheLeuLeuLeuTrpGlyValThr...TrpGly 16
::: |||||
353 ATTTCCTTTTACTACTTGGGGGATTAACGTGGTGGGT 315
seq_name: gb_est1:BE546037

seq_documentation_block:
LOCUS BE546037 443 bp mRNA EST 09-AUG-2000
DEFINITION 601072427P1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3458310 5',
mRNA sequence.
ACCESSION BE546037
VERSION BE546037.1 GI:9774682
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 443)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM849 row: f column: 07
High quality sequence stop: 441.
Location/Qualifiers

FEATURES
source

1..443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3458310"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: PCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 89 a 94 c 101 g 159 t
ORIGIN

alignment_scores:

Quality: 51.50 Length: 13
Ratio: 4.682 Gaps: 1
Percent Similarity: 84.615 Percent Identity: 76.923

alignment_block:

US-09-471-276-831_COPY_1_16 x BE546037 ..

Align seg 1/1 to: BE546037 from: 1 to: 443

5 ValValPheLeuLeuLeuTrpGlyValThr...TrpGly 16
||| |||||||||||||||||:||||| |||||
226 GTTCCCTTTTACTACTTGGGGGATACGCTGGGGG 264

seq_name: gb_est1:AI623983

seq_documentation_block:

LOCUS AI623983 467 bp mRNA EST 14-DEC-1999
DEFINITION ts25b11.x1 NCI-CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2229597 3',
mRNA sequence.
ACCESSION AI623983
VERSION AI623983.1 GI:4648914
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 467)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 1949 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 407
POLYA-No.

FEATURES
source

Location/Qualifiers
1..467
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2229597"
/clone_lib="NCI-CGAP_Pan1"
/tissue_type="adenocarcinoma"

/lab_host="DH10B"
/note="Organ: Pancreas; Vector: PCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"

BASE COUNT 170 a 101 c 100 g 96 t
ORIGIN

alignment_scores:

Quality: 51.50 Length: 13
Ratio: 4.682 Gaps: 1
Percent Similarity: 84.615 Percent Identity: 76.923

alignment_block:

US-09-471-276-831_COPY_1_16 x AI623983/rev ..

Align seg 1/1 to reverse of: AI623983 from: 1 to: 467

5 ValValPheLeuLeuLeuTrpGlyValThr...TrpGly 16
||| |||||||||||||||||:||||| |||||
235 GTTCCCTTTTACTACTTGGGGGATACGCTGGGGG 197

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2002, 16:48:23 ; Search time 19.35 Seconds

(without alignments)
496.020 Million cell updates/sec

Title: US-09-471-276-831

Perfect score: 661

Sequence: 1 MSMLVFLFLMGVWGPVTE.....LLELTGPKVLACSLADGAS 126

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	325	49.2	474	1 OMHUB	alpha-1-B-glycoprotein
2	78	11.8	192	2 S42940	viral infectivity
3	76	11.5	1327	2 T09402	immunoglobulin-lik
4	75.5	11.4	635	2 JCS896	killer cell inhibi
5	74.5	11.3	680	2 JCS895	killer cell inhibi
6	73	11.0	663	1 S46492	gelatinase A (EC 3
7	72.5	11.0	192	2 S42994	viral infectivity
8	71.5	10.8	489	2 A83033	probable aldehyde
9	71	10.7	336	2 I48471	Fc gamma (19G) rec
10	71	10.7	404	2 A46480	killer cell inhibi
11	69.5	10.5	841	2 JCS894	killer cell inhibi
12	69	10.4	192	2 S42950	viral infectivity
13	69	10.4	192	2 S42981	viral infectivity
14	69	10.4	192	2 S42967	viral infectivity
15	68.5	10.4	244	2 T04212	osmotic precursor
16	68.5	10.4	477	2 S65684	protoporphyrinogen
17	68	10.3	192	2 S42996	viral infectivity
18	68	10.3	192	2 S42947	viral infectivity
19	68	10.3	192	2 S33981	vif protein - huma
20	67.5	10.2	141	2 T14785	hypothetical prote
21	67.5	10.2	177	2 JCS748	coronafacic acid s
22	67.5	10.2	436	2 T36320	hypothetical prote
23	67.5	10.2	672	2 E84112	beta-galactosidase
24	67	10.1	192	2 S42999	viral infectivity
25	67	10.1	192	2 S42955	viral infectivity
26	67	10.1	192	2 S42953	viral infectivity
27	67	10.1	192	2 S42949	viral infectivity
28	67	10.1	362	2 T56130	HLA-B*5401 - huma
29	67	10.1	533	2 T38644	conserved hypothet

30	66.5	10.1	443	2 T31441	probable glutamyl-
31	66	10.0	192	2 S42997	viral infectivity
32	66	10.0	192	2 S42975	viral infectivity
33	66	10.0	192	2 S42954	viral infectivity
34	66	10.0	192	2 S43006	viral infectivity
35	66	10.0	192	2 S42972	viral infectivity
36	65.5	9.9	256	2 S34245	hypothetical prote
37	65	9.8	192	2 T08442	vif protein - huma
38	65	9.8	192	2 S42998	viral infectivity
39	65	9.8	192	2 S42959	viral infectivity
40	65	9.8	250	2 T08071	L-ascorbate peroxi
41	65	9.8	286	2 T49369	hypothetical prote
42	65	9.8	307	2 T07064	probable blac - My
43	65	9.8	455	2 B35916	site-specific reco
44	65	9.8	483	1 W2BE20	gene 20 protein -
45	64.5	9.8	258	1 S39747	lysin protein - Bac
46	64.5	9.8	745	2 B83834	hypothetical prote
47	64	9.7	192	2 S42942	viral infectivity
48	64	9.7	192	2 S42977	viral infectivity
49	64	9.7	192	2 S43001	viral infectivity
50	64	9.7	192	2 S42970	viral infectivity
51	64	9.7	192	2 S42992	viral infectivity
52	64	9.7	192	2 T01669	vif protein - huma
53	64	9.7	339	2 JCS7509	glycoprotein VI-1
54	64	9.7	362	2 JH0290	class I histocompa
55	64	9.7	1008	2 T04462	hypothetical prote
56	64	9.7	1069	2 D85383	hypothetical prote
57	63.5	9.6	258	2 B61228	collagen alpha 1(I
58	63.5	9.6	362	2 JH0288	class I histocompa
59	63.5	9.6	604	2 T08302	hypothetical prote
60	63.5	9.6	1669	1 CCH04B	collagen alpha 1(I
61	63	9.5	192	2 S42991	viral infectivity
62	63	9.5	192	2 S42968	viral infectivity
63	63	9.5	192	2 S42973	viral infectivity
64	63	9.5	192	2 S43004	viral infectivity
65	63	9.5	192	2 S42990	viral infectivity
66	63	9.5	192	2 S42969	viral infectivity
67	63	9.5	192	2 S42941	viral infectivity
68	63	9.5	330	2 T46256	brevican - human
69	63	9.5	366	2 I81232	lymphocyte antigen
70	63	9.5	366	2 I37135	MHC class I histoc
71	63	9.5	752	2 A65089	DNA topoisomerase
72	63	9.5	752	2 G85961	DNA topoisomerase
73	63	9.5	1090	2 S59077	cellulose 1,4-beta
74	62.5	9.5	347	2 G72684	hypothetical prote
75	62.5	9.5	356	2 JH0289	class I histocompa
76	62.5	9.5	508	1 JCS713	25-hydroxyvitamin
77	62.5	9.5	604	2 T08222	hypothetical prote
78	62.5	9.5	705	2 T31261	hypothetical prote
79	62.5	9.5	779	2 T14845	antifreeze-like pr
80	62.5	9.5	1188	2 S48861	gene ei protein -
81	62.5	9.5	1669	1 CCM548	collagen alpha 1(I
82	62	9.4	192	2 S43000	viral infectivity
83	62	9.4	192	2 S43003	viral infectivity
84	62	9.4	192	2 S42955	viral infectivity
85	62	9.4	316	2 T10436	hypothetical prote
86	62	9.4	362	2 T61907	probable transpos
87	62	9.4	366	2 T61866	MHC class I histoc
88	62	9.4	620	2 JH0821	95k golgi antigen
89	62	9.4	946	2 S71168	Ca2+-transporting
90	62	9.4	1020	2 D86402	protein envelope C
91	62	9.4	1020	2 T51925	Ca2+-transporting
92	62	9.4	1020	2 T51926	protein-tyrosine-P
93	61.5	9.3	160	2 F83329	probable sigma-70
94	61.5	9.3	181	2 A83626	FC gamma 2 recept
95	61.5	9.3	264	2 T46020	hypothetical prote
96	61.5	9.3	286	2 T11780	probable phosphor
97	61.5	9.3	438	2 S27848	enamelin precursor
98	61.5	9.3	1142	2 T37455	hypothetical prote
99	61.5	9.3	1167	2 T42204	filamentous hemag
100	61.5	9.3	3591	1 S21010	

C:Species: Mus musculus (house mouse)
C:Date: 18-Mar-1998 #sequence_revision 18-Mar-1998 #text_change 05-Nov-1999
C:Accession: J05895
R:Yamashita, Y.; Fukuta, D.; Tsuji, A.; Nagabukuro, A.; Matsuda, Y.; Nishikawa, Y.; Ohya
J. Biochem. 123, 358-368, 1998
A:Title: Genomic structures and chromosomal location of p91, a novel murine regulatory r
A:Reference number: J05894; MUID:98218758
A:Accession: J05895
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-680 <YAK>
A:Cross-references: GB:AF041035; NID:g2791691; PIDN:AA896927.1; PID:g2791692
C:Comment: This protein function as inhibitory cell-surface molecule against cell activa
C:Genetics:
A:Map position: 7
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-118/118-220, 221-315, 316-418, 419-517, 518-618/Domain: extracellular Ig-like #status F
F:616-674/Domain: transmembrane #status predicted <TMM>
F:675-680/Domain: cytoplasmic #status predicted <CYT>

Query Match 11.3%; Score 74.5; DB 2; Length 680;
Best Local Similarity 28.6%; Pred. No. 4.9;
Matches 32; Conservative 11; Mismatches 48; Indels 21; Gaps 4;

OY 6 VELLMGVWGPVTEA---AIFETQXSLMAESENHLTKLCCDADVPDGPDSRLP 61
DB 215 VLLVSGNLQKPTIKPESVITSRAMITWCO-----GNLDEV-----YFLHNE 260

OY 62 QEWGAQEPVHLDSPAIKHOFLLTGDTPQ---GRYCRSGSLSTGWXQSLKLEL 110
DB 261 KSGKQSTGTTLQDGPGRKGRFPISVTQGHAGQRCYCSAGMSQPSDLLEL 312

RESULT 6
S46492
gelatinase A (EC 3.4.24.24) precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S46492
R:Aimes, R.T.; French, D.L.; Outgley, J.P.
Biochem. J. 300, 729-736, 1994
A:Title: Cloning of a 72 kDa matrix metalloproteinase (gelatinase) from chicken embryo f
A:Reference number: S46492; MUID:94280397
A:Accession: S46492
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-663 <AIM>
A:Cross-references: EMBL:U07775; NID:g504475; PIDN:AA19596.1; PID:g504476
A:Note: In the authors' translation 205-Asp is shown after residue 201 and, consequently
C:Superfamily: gelatinase A; fibronectin type II repeat homology; hemopexin repeat homol
C:Keywords: hydrolase; metalloproteinase; zinc; zymogen
F:67-216, 391-443/Domain: matrix metalloproteinase homology #status atypical <MMP>
F:230-271/Domain: fibronectin type II repeat homology <F1>
F:288-329/Domain: fibronectin type II repeat homology <F8>
F:346-387/Domain: fibronectin type II repeat homology <F9>
F:466-663/Domain: hemopexin repeat homology <FXN>
F:399,400,404,410/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
F:400,404,410/Binding site: zinc, catalytic (His) (active) #status predicted
F:401/Active site: Glu #status predicted

Query Match 11.0%; Score 73; DB 1; Length 663;
Best Local Similarity 23.6%; Pred. No. 6.9;
Matches 26; Conservative 13; Mismatches 37; Indels 34; Gaps 4;

OY 17 PVTEAIFETQXSLMAESENHLTKLCCDADVPDGPDSRLP 69
DB 530 PDEKAVF-----AGNYWYTTASNLDRGPKKLTSLGLPDDVDRIDAAFNWGRN-- 580

OY 70 VHLDSPAIKHOFLLTGDTPGRYCRSGSLSTGWXQSLKLELTPKVLACS 119
DB 581 -----KTYIFSGDRYWKYN-----EKKKKMELATPKFIADS 612

RESULT 7
S42994
viral infectivity factor vif - human immunodeficiency virus type 1
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Sep-1999
C:Accession: S42994
R:Wieland, U.; Hartmann, J.; Subr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.
submitted to the EMBL Data Library, March 1994
A:Description: In vivo genetic variability of the HIV-1 gene.
A:Accession number: S42940
A:Accession: S42994
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <WIE>
A:Cross-references: EMBL:Z30684; NID:g459605; PIDN:CA83164.1; PID:g459606
C:Superfamily: AIDS vif protein

Query Match 11.0%; Score 72.5; DB 2; Length 192;
Best Local Similarity 25.6%; Pred. No. 1.9;
Matches 20; Conservative 12; Mismatches 37; Indels 9; Gaps 2;

OY 8 LLLMGV-----TWGPVTEAIFETQXSLMAESENHLTKLCCDADVPDGPDSRLP 61
DB 8 MIVMOVDRMRIRITWKSILVHNHIVYSKAKGWFYRHHYESTHRKISSEVHPIGLDARLVIT 67

OY 62 QEWG---AQEPVHLSPA 76
DB 68 TYWGLNTGERDNHLDGCA 85

RESULT 8
A83033
probable aldehyde dehydrogenase PA4899 [imported] - Pseudomonas aeruginosa (strain PA
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83033
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: A83033
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-489 <STO>
A:Cross-references: GB:AE004903; GB:AE004901; NID:g9951173; PIDN:AA08284.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4899
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 10.8%; Score 71.5; DB 2; Length 489;
Best Local Similarity 27.4%; Pred. No. 7.1;
Matches 32; Conservative 16; Mismatches 54; Indels 15; Gaps 4;

OY 16 GPVTEAIFETQXSLMAESENHLTKLCCDADVPDGPDSRLP 74
DB 67 GPSARAVALTKAVEVDRRHEELVDMIIIR-----ESGSTRKALCLENGAAARAITLESA 119

OY 75 --PAIKHOFLLTGDTPGR---YCRSGSLSTGWXQSLKLELTPKVLACS 125
DB 120 SPFARVHGRIVESDVPKSKRYRSALIGVGVISPNMFLPLHT--QSTIAPALALGMA 175

RESULT 9
I48471
Fc gamma (IgG) receptor high affinity - mouse
C:Species: Mus musculus (house mouse)

RESULT 11

RESULT 13

viral infectivity factor vif - human immunodeficiency virus type 1
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Sep-1999

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36320
R:Oliver, K.; Harris, D.; James, K.D.; Parthill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z21575
A:Accession: T36320
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-436
A:Cross-references: EMBL:AL049841; PIDN:CA842747.1; GSPDB:GN00070; SCOEDB:SCE9.01
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE9.01

Query Match 10.2%; Score 67.5; DB 2; Length 436;
Best Local Similarity 26.8%; Pred. No. 17;
Matches 30; Conservative 11; Mismatches 50; Indels 21; Gaps 5;

QY 3 MLVVFLLMGVTPGPTFAI---FYETQXSLMASEHXLTGCGDADVPDPPGDSRL 58
DB 179 VLAVALTLTGATWGLAAGVPTLMLATHTVLTALPLVGAAGASSAG-----RP 233
QY 59 PAVQENG-AQEPVHLDSPAIKHQFLTGTGRCRSGISTGMXLSKLE 109
DB 234 VGEYDMTVAESF---DGLRIDHGL-----DRAHETVPPRGVQTRLIVE 274

RESULT 23
E84112
beta-galactosidase BH3701 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: E84112
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20263314
A:Accession: E84112
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-672 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA807420.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3701
C:Superfamily: Bacillus beta-galactosidase

Query Match 10.2%; Score 67.5; DB 2; Length 672;
Best Local Similarity 31.4%; Pred. No. 28;
Matches 27; Conservative 9; Mismatches 29; Indels 21; Gaps 5;

QY 35 SEHXLTGCGDADVPDPPGDSRLPAVQENGAEVHLDSPAIKHQFLTGTG---QGRY 91
DB 492 NENDIVTIG-----GYPELR-DLGLTW--VEEIDALPPEKKNQIVITNDIGSLTGY 541
QY 92 RCRSGISTGMXLSKLELTGPVLA 117
DB 542 ECR-----LTFDIIHSECAVDLA 559

RESULT 24
S42999
viral infectivity factor vif - human immunodeficiency virus type 1
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Sep-1999
C:Accession: S42999
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.
submitted to the EMBL Data Library, March 1994
A:Description: In vivo genetic variability of the HIV-1 gene.
A:Reference number: S42940
A:Accession: S42999

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <WIE>
A:Cross-references: EMBL:Z30690; NID:g459617; PIDN:CAA83170.1; PID:g459618
C:Superfamily: AIDS vif protein

Query Match 10.1%; Score 67; DB 2; Length 192;
Best Local Similarity 20.0%; Pred. No. 7.3;
Matches 20; Conservative 17; Mismatches 39; Indels 24; Gaps 3;

QY 8 LLLMGV-----TWGPTFAIIFYETQXSLMASEHXLTGCGDADVPDPPGDSRLPAV 61
DB 8 MIVQVDRMRIRKTWNSLVKHHMTVSGAKGWSYRHNYESTNPRISSEVHPIGLDARLYVT 67
QY 62 QEWGAEVHLDSPAIKHQFLTGTGRCRSGISTGM 101
DB 68 TYWG-----LHTGERD--WHLGQGVSIEM 89

RESULT 25
S42995
viral infectivity factor vif - human immunodeficiency virus type 1
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Sep-1999
C:Accession: S42995
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.
submitted to the EMBL Data Library, March 1994
A:Description: In vivo genetic variability of the HIV-1 gene.
A:Reference number: S42940
A:Accession: S42995
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <WIE>
A:Cross-references: EMBL:Z30685; NID:g459607; PIDN:CAA83165.1; PID:g459608
C:Superfamily: AIDS vif protein

Query Match 10.1%; Score 67; DB 2; Length 192;
Best Local Similarity 21.0%; Pred. No. 7.3;
Matches 21; Conservative 15; Mismatches 40; Indels 24; Gaps 3;

QY 8 LLLMGV-----TWGPTFAIIFYETQXSLMASEHXLTGCGDADVPDPPGDSRLPAV 61
DB 8 MIVQVDRMRIRKTWNSLVKHHMTVSGAKGWSYRHNYESTNPRISSEVHPIGLDARLYVT 67
QY 62 QEWGAEVHLDSPAIKHQFLTGTGRCRSGISTGM 101
DB 68 TYWG-----LHTGERD--WHLGQGVSIEM 89

RESULT 26
S42953
viral infectivity factor vif - human immunodeficiency virus type 1
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Sep-1999
C:Accession: S42953
R:Wieland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.
submitted to the EMBL Data Library, March 1994
A:Description: In vivo genetic variability of the HIV-1 gene.
A:Reference number: S42940
A:Accession: S42953
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <WIE>
A:Cross-references: EMBL:Z30614; NID:g459470; PIDN:CAA83091.1; PID:g459471
C:Superfamily: AIDS vif protein

Query Match 10.1%; Score 67; DB 2; Length 192;
Best Local Similarity 20.0%; Pred. No. 7.3;
Matches 20; Conservative 16; Mismatches 40; Indels 24; Gaps 3;

QY	94	RSG	LG	ST	GM	XQ	103
		:	:	:	:	:	
Db	134	SNG	IV	NT	WF	Q	143

RESULT 31

viral infectivity factor vif - human immunodeficiency virus type 1
 C:Species: human immunodeficiency virus type 1, HIV-1
 C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Sep-1999
 C:Accession: S42997
 R:Wieland, U.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.
 submitted to the EMBL Data Library, March 1994
 A:Description: In vivo genetic variability of the HIV-1 gene.
 A:Reference number: S42940
 A:Accession: S42997
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-192 <WIE>
 A:Cross-references: EMBL:Z30688; NID:g459613; PIDN:CA83f68.1; PID:g459614
 C:Superfamily: AIDS vif protein

Query Match	10.0%;	Score 66;	DB 2;	Length 192;
Best Local Similarity	21.0%;	Pred. NO. 9.3;		
Matches	21;	Conservative	16;	Mismatches 39;
			Indels	24;
			Gaps	3

```

Qy      8 LLLKGV-----TWGPVTEAIFETQXSLMAESEHXLKTLGGCDADVBPQGDSLPAV 61
      ::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      8 MIVQVDRLRIKTIWKSLLVKNHMYISGKAKGFFYRNHRETHPRISSEVNIPLGDARLVIT 67

```

```

OY      62 QEMGAQEPVHLDSPAIRKQFLLTGDTGGRKCRSGSLGTCW 101
          ||      |||:      |||:
DB      68 TYWG-----LRTGERD--WHLGQGVSIEM 89

```

RESULT 32

S42975
viral infectivity factor vif - human immunodeficiency virus type 1
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Sep-1999
C:Accession: S42975
R:Meland, U.; Hartmann, J.; Shih, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.
submitted to the EMBL Data Library, March 1994
A:Description: In vivo genetic variability of the HIV-1 gene.
A:Reference number: S42940
A:Accession: S42975
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <MT>
A:Cross-references: EMBL:Z30636; NID:9459514; PIDN:CAAB83113.1; PID:9459515
C:Superfamily: AIDS vif protein

Query Match	10.0%;	Score 66;	DB 2;	Length 192;
Best Local Similarity	21.0%;	Pred. No. 9.3;		
Matches	21;	Conservative	15;	Mismatches 40;
			Indels	24;
			Gaps	3

```
Oy      8 LLLMGV-----TWGPTTAAIFETQXSLMAESENHKLKTGGCDADVPGRPGDSRLPAV 67
        ::::| |:::| |:::| |:::| |:::|
Db      8 MIVWQDRMRIRIPWNLSLVKHHMYSGKARGWTYRHHYESTNPRVSSEVHIPLGDARLVT 67
```

Qy 62 QENGAQEPVHLSDPAIKHOFLLTGDTQCRKCSGLSTGW 101
 || | | |
 Db 68 TYWG-----LHNGERD--WHLLGQVSLEW 89

RESULT 33

viral infectivity factor vif - human immunodeficiency virus type 1
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Sep-1999

C:Accession: 542954
 R:Meland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.
 submitted to the EMBL data library, March 1994
 A:Description: In vivo genetic variability of the HIV-1 gene.
 :Reference number: 542940

Query Match	10.0%;	Score 66;	DB 2;	Length 192;
Best Local Similarity	20.0%;	Pred. No. 9.3;		
Matches	20;	Conservative 17;	Mismatches 39;	Indels 24;
			Gaps	3;

```
Oy      8 LLLMGV-----TWGPIVEAIFETQXSLMAESEHXLKTLCGDADVPGRPGDSRLPAV   61  
        ::::| | : : : : | | : : : : ||  
  
Db      8 MIVWQDRMRIRFTWKLSLVKHHMYVSCKAKGVYKKHNESTHPRISSFEVHIPGEARLVII   67
```

```

QY      62 QENGAQDEPVHLDSPAIKHQLLTGDTQGNKRSGLSTGW 101
          ||      | || : || |
DB      68 TYWG-----LHNGERD--WHLGGGVSIEM 89

```

RESULT 34

viral infectivity factor vif - human immunodeficiency virus type 1
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Sep-1999
C:Accession: SA3006; SA2948
R:Wittland, U.; Hartmann, J.; Sühr, H.; Salzberger, B.; Eggers, H.J.; Kuehn, J.E.
submitted to the EMBL Data Library, March 1994
A:Description: In vivo genetic variability of the HIV-1 gene.
A:Reference number: SA2940
A:Accession: SA3006
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <MIE>
A:Cross-references: EMBL:Z30661; NTD:g459631; P1DN:CAH83133.1; P1D:g459632; EMBL:Z30661
C:Superfamily: AIDS vif protein

Query Match	10.0%;	Score 66;	DB 2;	Length 192;
Best Local Similarity	21.0%;	Pred. No. 9.3;		
Matches	21;	Conservative	15;	Mismatches 40;
			Indels	24;
			Gaps	3

```

Oy      8 LLLMGV-----TWGPRVTEAIFYETQXSLAESEHNKLTGGCDADVPGRPGDSRLP 61
      ::: | | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db      8 MIVQVDNRMRIRFWKSLVKNHMYISGAKRMFYRNHNHSTHPRISSEVNHIRLGDALVIT 67

```

```

Oy      62 OEMGAQEPVHLDSPAIKHQLFTEDTGKRYKRCRSGLSICW 10
          ||      | || :      | : | |
Db      68 TYNG-----LHTERD--WHLGGGVSTEW 89

```

MEBODI
S42972

viral infectivity factor vif - human immunodeficiency virus type 1
 C:Species: human immunodeficiency virus type 1, HIV-1
 C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Sep-1999
 C:Accession: S42972
 R:Wietland, U.; Hartmann, J.; Suhr, H.; Salzberger, B.; Eggers, H.J.; Koehn, J.E.
 submitted to the EMBL Data Library, March 1994
 A:Description: In vivo genetic variability of the HIV-1 gene.
 A:Reference number: S42940
 A:Accession: S42972
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-192 <MID>
 A:Cross-references: EMBL:Z30633; MID:g459508; PIDN:CAA83110.1; PID:g459509
 C:Superfamily: AIDS vif protein

Query Match 10.0%; Score 66; DB 2; Length 192;
Best Local Similarity 20.0%; Pred. NO. 9.3;
Matches 20; Conservative 17; Mismatches 39; Indels 24; Gaps 3;

A:Reference number: Z16333
A:Accession: T08071
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-250 <TAY>
A:Cross-references: EMBL:AF038839; NID:92746726; PIDN:AAB94927.1; PID:92746727
C:Superfamily: cytochrome-c peroxidase
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
P:42/Active site: His (distal axial ligand) *status predicted
P:163/Binding site: heme iron (His) (proximal axial ligand) *status predicted
P:179,208/Active site: Trp, Asp *status predicted

Query Match 9.8%; Score 65; DB 2; Length 250;
Best Local Similarity 29.8%; Pred. No. 16;
Matches 28; Conservative 5; Mismatches 41; Indels 20; Gaps 5;

QY 48 DVPGPDSRLPVDQEMGAQEPVHLDSPAIRKQFLITG-----TOGRYRCR-SGLS 98
| | | | | : | | | | | : | | | | |
DB 121 DKRPQPPPEGRLPDARK-GCD---HLRQVLLKQMLTDDIIVALSQAHTLGRYRAAPSQFE 176
| | | | | : | | | | | : | | | | |
QY 99 TGMXOLSKLE-----LTGPRVLAQSLALDCA 125
| | | | | : | | | | | : | | | | |
DB 177 GAWTSNPLIFDMSYFELLTGEKGLQLVSDKA 210
| | | | | : | | | | | : | | | | |

RESULT 41
T49369
hypothetical protein B1D1.230 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T49369
R:Schulte, U.; Aign, V.; Hohsels, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49369
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-286 <SCH>
A:Cross-references: EMBL:AL355927; GSPDB:GN00116; NCSP:B1D1.230
A:Experimental source: BAC clone B1D1; strain OR74A
C:Genetics:
A:Gene: NCSP:B1D1.230
A:Map position: 6
A:Introns: 65/2, 159/1
C:Superfamily: Neurospora crassa hypothetical protein B1D1.230

Query Match 9.8%; Score 65; DB 2; Length 286;
Best Local Similarity 33.8%; Pred. No. 19;
Matches 22; Conservative 5; Mismatches 28; Indels 10; Gaps 3;

QY 27 TOXSLMAESEXLKTGCGCDADVPYPPGDSRLPVA-----QEWG---AOEPVHLDSPAIRK 78
| | | | | : | | | | | : | | | | |
DB 106 TOLRLVVVSAQRAKADG--GVEVPGRTGOSRYPVATHCDLHSGMAPRPPQPAHSKVSAYR 163
| | | | | : | | | | | : | | | | |
QY 79 HQFL 83
| | | | | : | | | | | : | | | | |
DB 164 WKVVL 168
| | | | | : | | | | | : | | | | |

RESULT 42
G70764
probable blac - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: G70764
R:Coale, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rulter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987
A:Accession: G70764
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-307 <COL>
A:Cross-references: GB:Z73966; GB:AL123456; NID:93261577; PIDN:CAA98216.1; PID:g13702
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: blac
C:Superfamily: beta-lactamase I

Query Match 9.8%; Score 65; DB 2; Length 307;
Best Local Similarity 27.9%; Pred. No. 21;
Matches 36; Conservative 17; Mismatches 44; Indels 32; Gaps 7;

QY 12 GVTWGPVTEAIFYE--TQXSL-----MAESEXLKTG---QCDADVP---- 50
| | | | | : | | | | | : | | | | |
DB 128 GMTIGOLCDAIRYSDGTANILLADLGGPGGTAFTGTLSLGDVSRDLAEEPELNR 187
| | | | | : | | | | | : | | | | |
QY 51 GPPGDSR---LPAVQEMGAQEPVHLDSPAIRKQFLITG-----DTGRYRCRSGLSGWXO 103
| | | | | : | | | | | : | | | | |
DB 188 DPGGDERDTTPPAIALVLQQLVLAGNALPPDRKALLTDMMARNTGAKRIAGFPADW-- 245
| | | | | : | | | | | : | | | | |
QY 104 LSKILELTG 112
| | | | | : | | | | | : | | | | |
DB 246 --KVIDKTG 252
| | | | | : | | | | | : | | | | |

RESULT 43
B36916
site-specific recombinase Int - Streptomyces lividans
C:Species: Streptomyces lividans
C:Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
C:Accession: B36916
R:Brasch, M.A.; Pettis, G.S.; Lee, S.C.; Cohen, S.N.
J. Bacteriol. 175, 3067-3074, 1993
A:Title: Localization and nucleotide sequences of genes mediating site-specific recom
A:Reference number: A36916; MUID:93259953
A:Accession: B36916
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-455 <BRA>
A:Note: sequence extracted from NCBI backbone (NCBIN:131954, NCBIF:131957)

Query Match 9.8%; Score 65; DB 2; Length 455;
Best Local Similarity 30.0%; Pred. No. 32;
Matches 21; Conservative 8; Mismatches 31; Indels 10; Gaps 2;

QY 5 VFLLLMGVY-WGPVTEAIFYETQXSL-----MAESEXLKTGCGCDADVPYPPG 54
| | | | | : | | | | | : | | | | |
DB 338 LVMLRLMHWYAYGVAPDGRILFTORGLIODTGYGEVMAEARSRLTPAOCASLAKRPY 397
| | | | | : | | | | | : | | | | |
QY 55 DSRLLPVOEW 64
| | | | | : | | | | | : | | | | |
DB 398 DLRRHAAYSTW 407
| | | | | : | | | | | : | | | | |

RESULT 44
WZRE20
gene 20 protein - human herpesvirus 3
C:Species: human herpesvirus 3, varicella-zoster virus
C:Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C:Accession: B27343
R:Davidson, A.J.; Scott, J.E.
J. Gen. Virol. 67, 1759-1816, 1986
A:Title: The complete DNA sequence of varicella-zoster virus.
A:Reference number: A27343; MUID:86306657
A:Accession: B27343
A:Molecule type: DNA
A:Residues: 1-483 <DAV>
A:Cross-references: EMBL:X04370; NID:g959989; PIDN:CA27903.1; PID:g60009

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2002, 16:45:13 ; Search time 19.31 Seconds
(without alignments)
146,837 Million cell updates/sec

Title: US-09-471-276-831

Perfect score: 661

Sequence: 1 MSMLVFLLMGVTWGPVTE.....LLELTGRVLACSLALDGAS 126

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 100 summaries

Database :

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTDUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87.5	13.2	631	4	US-09-345-468-12
2	75.5	11.4	244	3	US-09-135-782-4
3	74	11.2	651	4	US-08-985-950-22
4	72.5	11.0	244	4	US-09-193-191-4
5	68.5	10.4	390	2	US-08-979-424-1
6	67	10.1	234	4	US-09-113-750A-38
7	66	10.0	415	1	US-08-999-774A-27
8	65.5	9.9	503	4	US-08-964-523B-27
9	65	9.8	501	4	US-08-857-076-102
10	65	9.8	685	2	US-08-878-989-1
11	65	9.8	685	3	US-09-136-282-2
12	65	9.8	685	4	US-09-272-796-1
13	65	9.8	685	4	US-09-505-744-2
14	64	9.7	249	4	US-09-345-468-5
15	64	9.7	319	4	US-09-345-468-5
16	64	9.7	339	4	US-09-345-468-3
17	63.5	9.6	431	4	US-08-985-950-14
18	63.5	9.6	431	4	US-08-985-950-20
19	63.5	9.6	615	4	US-08-985-950-16
20	63.5	9.6	615	4	US-08-985-950-18
21	62.5	9.5	385	4	US-09-071-224-19
22	62.5	9.5	389	4	US-09-071-224-28
23	62.5	9.5	405	3	US-09-012-072-2
24	62.5	9.5	405	4	US-09-120-601-2
25	62.5	9.5	425	4	US-09-071-224-6
26	62.5	9.5	508	4	US-09-111-730-2
27	61.5	9.3	303	4	US-09-071-224-23

28	61.5	9.3	389	4	US-09-071-224-22	Sequence 22, Appl
29	61.5	9.3	389	4	US-09-071-224-24	Sequence 24, Appl
30	61.5	9.3	389	4	US-09-071-224-25	Sequence 25, Appl
31	61.5	9.3	389	4	US-09-071-224-27	Sequence 27, Appl
32	61.5	9.3	392	4	US-09-071-224-18	Sequence 18, Appl
33	61.5	9.3	425	4	US-09-071-224-4	Sequence 4, Appl
34	61.5	9.3	1248	2	US-08-348-353-17	Sequence 17, Appl
35	61.5	9.3	1248	2	US-08-465-965-17	Sequence 17, Appl
36	61.5	9.3	1248	3	US-08-465-966-17	Sequence 17, Appl
37	61	9.2	3567	2	US-07-642-734C-4	Sequence 4, Appl
38	61	9.2	3567	3	US-08-439-009A-4	Sequence 4, Appl
39	60.5	9.2	389	4	US-09-071-224-26	Sequence 26, Appl
40	60.5	9.2	384	2	US-08-449-645A-19	Sequence 19, Appl
41	60.5	9.2	984	2	US-08-702-367A-19	Sequence 19, Appl
42	60.5	9.2	984	5	PCT-US95-04681-19	Sequence 19, Appl
43	60	9.1	254	2	US-08-667-939A-2	Sequence 2, Appl
44	60	9.1	421	4	US-08-759-628-5	Sequence 5, Appl
45	60	9.1	553	1	US-07-683-957B-3	Sequence 3, Appl
46	60	9.1	912	5	PCT-US95-03747-2	Sequence 2, Appl
47	60	9.1	3816	4	US-09-428-517-3	Sequence 3, Appl
48	59.5	9.0	389	4	US-09-071-224-30	Sequence 30, Appl
49	59.5	9.0	1694	1	US-08-494-168-2	Sequence 2, Appl
50	59	8.9	389	4	US-09-071-224-26	Sequence 26, Appl
51	59	8.9	389	4	US-09-071-224-31	Sequence 31, Appl
52	59	8.9	501	3	US-08-906-791-2	Sequence 2, Appl
53	59	8.9	501	4	US-09-111-730-1	Sequence 1, Appl
54	59	8.9	509	2	US-08-845-566-1	Sequence 1, Appl
55	59	8.9	600	2	US-08-370-156-4	Sequence 1, Appl
56	59	8.9	600	3	US-08-814-095-4	Sequence 4, Appl
57	59	8.9	600	4	US-08-975-084-1	Sequence 1, Appl
58	58.5	8.9	106	3	US-09-083-351-9	Sequence 9, Appl
59	58.5	8.9	106	4	US-09-083-352-9	Sequence 9, Appl
60	58.5	8.9	378	2	US-08-986-217-6	Sequence 6, Appl
61	58.5	8.9	1241	4	US-09-040-774-2	Sequence 6, Appl
62	58	8.8	147	2	US-08-469-587A-15	Sequence 15, Appl
63	58	8.8	147	3	US-08-318-038D-14	Sequence 14, Appl
64	58	8.8	701	4	US-08-227-496C-18	Sequence 18, Appl
65	58	8.8	3729	2	US-09-370-368-10	Sequence 10, Appl
66	58	8.8	3729	2	US-08-804-227C-4	Sequence 4, Appl
67	57.5	8.7	303	4	US-09-071-224-21	Sequence 21, Appl
68	57.5	8.7	385	4	US-09-071-224-20	Sequence 20, Appl
69	57.5	8.7	388	4	US-09-071-224-17	Sequence 17, Appl
70	57.5	8.7	422	4	US-09-071-224-2	Sequence 2, Appl
71	57.5	8.7	3491	2	US-07-642-734C-2	Sequence 2, Appl
72	57.5	8.7	3491	3	US-08-439-009A-2	Sequence 2, Appl
73	57	8.6	139	4	US-08-444-818-174	Sequence 174, App
74	57	8.6	192	4	US-09-124-900-4	Sequence 4, Appl
75	57	8.6	203	3	US-08-463-210-10	Sequence 10, Appl
76	57	8.6	233	2	US-08-667-939A-8	Sequence 8, Appl
77	57	8.6	254	2	US-08-667-939A-3	Sequence 3, Appl
78	57	8.6	441	1	US-08-188-281B-21	Sequence 21, Appl
79	57	8.6	441	1	PCT-US94-07280-21	Sequence 21, Appl
80	57	8.6	441	5	PCT-US95-01087-21	Sequence 21, Appl
81	57	8.6	453	1	US-08-188-281B-16	Sequence 16, Appl
82	57	8.6	453	5	PCT-US94-07280-16	Sequence 16, Appl
83	57	8.6	453	5	PCT-US95-01087-16	Sequence 16, Appl
84	57	8.6	453	6	5284931-3	Patent No. 5284931
85	57	8.6	490	1	US-08-188-281B-15	Sequence 15, Appl
86	57	8.6	490	5	PCT-US94-07280-15	Sequence 15, Appl
87	57	8.6	490	5	PCT-US95-01087-15	Sequence 15, Appl
88	57	8.6	505	3	US-08-318-039A-1	Sequence 1, Appl
89	57	8.6	505	3	US-08-318-038D-15	Sequence 15, Appl
90	57	8.6	505	4	US-08-327-496C-19	Sequence 19, Appl
91	57	8.6	505	4	US-08-435-568A-1	Sequence 1, Appl
92	57	8.6	505	4	US-09-240-915-6	Sequence 1, Appl
93	57	8.6	505	4	US-09-591-435-6	Sequence 6, Appl
94	57	8.6	507	2	US-08-469-587A-14	Sequence 14, Appl
95	57	8.6	507	2	US-08-469-587A-14	Sequence 14, Appl
96	57	8.6	531	2	US-08-789-078-3	Sequence 3, Appl
97	57	8.6	531	2	US-08-752-633-3	Sequence 3, Appl
98	57	8.6	531	5	PCT-US95-04886-3	Sequence 3, Appl
99	57	8.6	532	1	US-07-618-286-1	Sequence 1, Appl
100	57	8.6	532	1	US-08-196-003-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-345-468-12
Sequence 12, Application US/09345468
Patent No. 6245527
GENERAL INFORMATION:
APPLICANT: Bustfield, S.
APPLICANT: Villaveal, J.
APPLICANT: Jandrot-Perrus, M.
APPLICANT: Valchenker, W.
TITLE OF INVENTION: GLYCOPROTEIN VI AND USES THEREOF
FILE REFERENCE: 7853-147
CURRENT APPLICATION NUMBER: US/09/345,468
CURRENT FILING DATE: 1999-06-30
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 631
TYPE: PRT
ORGANISM: Homo sapiens
US-09-345-468-12

Query Match 13.2%; Score 87.5; DB 4; Length 631;
Best Local Similarity 27.5%; Pred. No. 0.0078;
Matches 38; Conservative 11; Mismatches 38; Indels 51; Gaps 6;

QY 1 MSMLVFLLMGVTWGPVTEAIFETQXSLMAESEHXLKTLGOCADVPGPDSRLPA 60
DB 1 MTPALTRALLGLSLGPRTHVQAGPPKPTLMAE-----FG-----S 37
QY 61 VQEWG-----AQE-----PVHLD-----SPAIRKQFLITGDTQ---GRYR 92
DB 38 VISMGSVVTIMCGSLAQEYRLDKESPEPLDRBNPLERKKNKARSIPSMTHMGRRYR 97
QY 93 CRSGLSTGMXQLSKLEL 110
DB 98 CHYYSAGWSEPSDPLEL 115

RESULT 2
US-09-135-782-4
Sequence 4, Application US/09135782
Patent No. 6027929
GENERAL INFORMATION:
APPLICANT: Xu, Shuang-Yong
TITLE OF INVENTION: Method For Cloning And Producing The Nspi Restriction
TITLE OF INVENTION: Endonuclease In E. coli And Purification Of The
FILE REFERENCE: NEB-143
CURRENT APPLICATION NUMBER: US/09/135,782
CURRENT FILING DATE: 1998-08-18
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 244
TYPE: PRT
ORGANISM: No. 6027929Ucc sp.
US-09-135-782-4

Query Match 11.4%; Score 75.5; DB 3; Length 244;
Best Local Similarity 39.3%; Pred. No. 0.076;
Matches 22; Conservative 5; Mismatches 18; Indels 11; Gaps 4;

QY 55 DSRUPAEOEGADPEVILDS---PAIKHOF-LITGDTQGRYR-----CRSGLSTG 100
DB 132 DHRFP-MERWGASEPRLTSMNDNEIKRKHQLKDTSGNHMLLKSNSCERCICKTG 186

RESULT 3

US-08-985-950-22
Sequence 22, Application US/08985950
Patent No. 6140076
GENERAL INFORMATION:
APPLICANT: Adema, Gosse Jan
TITLE OF INVENTION: Isolated Mammalian Monocyte Cell Genes;
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESS: DMAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,950
FILING DATE: 05-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/041,279
FILING DATE: 21-MARCH-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/033,181
FILING DATE: 16-DEC-1996
APPLICATION NUMBER: US 60/032,252
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0670K
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)852-9196
TELEFAX: (650)496-1204
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 651 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-985-950-22

Query Match 11.2%; Score 74; DB 4; Length 651;
Best Local Similarity 26.6%; Pred. No. 0.47;
Matches 34; Conservative 14; Mismatches 50; Indels 30; Gaps 5;

QY 1 MSMLVFLLMGVTWGPVTEAIFETQXSLMAESEHXLKTLGOCADVPGPDSRLPA 60
DB 1 MTPILVTLICGLSLGPRTHVQAGHLPKPTLMAEPGSI-----TQGSFVTLRCGG 51
QY 61 VQE-----WGAQEPVHLDSPAIKHOF---LITGDTQGRYRCRSGLSLST-GMX 102
DB 52 GQETQEVRLYREKKTA PWLIRIPOEL---VKKGQFPIPSITWHAQRYRCYCGSDTAGRS 108
QY 103 QLSKLEL 110
DB 109 ESSDPLEL 116

RESULT 4
US-09-193-191-4
Sequence 4, Application US/09193191
Patent No. 6130078
GENERAL INFORMATION:
APPLICANT: XIAO, JIAN-PING

APPLICANT: XU, SHUANG-YONG
TITLE OF INVENTION: METHOD FOR CLONING THE NSPHI RESTRICTION-MODIFICATION
TITLE OF INVENTION: SYSTEM IN E. COLI AND PRODUCING THE RECOMBINANT NSPHI
TITLE OF INVENTION: RESTRICTION ENDONUCLEASE
FILE REFERENCE: NSPHI
CURRENT APPLICATION NUMBER: US/09/193,191
CURRENT FILING DATE: 1998-11-17
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 4
LENGTH: 244
TYPE: PRT
ORGANISM: No. 6130078tcc sp.
US-09-193-191-4

Query Match 11.0%; Score 72.5; DB 4; Length 244;
Best Local Similarity 37.5%; Pred. No. 0.19;
Matches 21; Conservative 6; Mismatches 18; Indels 11; Gaps 4;

OY 55 DSRPAVOEMGAOEPVHLDSPA---IKHOF-LLTGPTGGRYR-----CRSGISTG*100
DB 132 DHRFP-MENMGASBPPLTMSDDEIKOKFOLLKDGASGNHNLRSRSCERCICKTG 186

RESULT 5

US-08-979-424-1
Sequence 1, Application US/08979424
Patent No. 5942606
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VIRAL RECEPTOR PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,424
FILING DATE: Filed Herewith.
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0405 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 390 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNCEFT03
CLONE: 1232054
US-08-979-424-1

Query Match 10.4%; Score 68.5; DB 2; Length 390;
Best Local Similarity 24.6%; Pred. No. 1.2;
Matches 29; Conservative 15; Mismatches 43; Indels 31; Gaps 4;
OY 37 HXIKTLGQCDADVPGBPGDRLPAVOEMGAOEPVHLDS---PAIKHOF-----81
DB 140 HSIKTL-ELNVLVPAPSPCRLOGVPHVGANVTLSCQSPRSKPAVOYQMDROLPSFOTFE 198

OY 82 -----LLTGTGGRYRCRSLSTGKXOLSKLLEL-TGPKYLACSLALDG 124
DB 199 APALDVIRGSLSTFNLSSMGAVYCKAHNEVGTACQNTLESTGPAAVAGAVG 256

RESULT 6

US-09-113-750A-38
Sequence 38, Application US/09113750A
Patent No. 6294176
GENERAL INFORMATION:
APPLICANT: David E. Junker and Mark D. Cochran
TITLE OF INVENTION: Recombinant Raccoonpox Virus
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESS: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/113,750A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 55744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)262-0400
TELEFAX: (212)664-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-113-750A-38

Query Match 10.1%; Score 67; DB 4; Length 254;
Best Local Similarity 29.9%; Pred. No. 1;
Matches 26; Conservative 6; Mismatches 25; Indels 30; Gaps 5;

OY 21 AAIFETQXSIMAESHXLTGQCDADVPG---PG---DSRLPAVOEM-----64
DB 63 AAIIYEPGRSILMGRIGH--DRGSDHDELGFVNPGLSSEGLHIVYAMLAFLSFSYTA 120

OY 65 -----GAOEPVHLDSPAIKHOF 82
DB 121 OFHPEIFGIGVSKYVVD---IKHOFI 144

RESULT 7

US-08-464-523B-27
Sequence 27, Application US/08464523B
Patent No. 5723761
GENERAL INFORMATION:

;; CURRENT FILING DATE: 1997-05-15
;; NUMBER OF SEQ ID NOS: 114
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 102
;; LENGTH: 501
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-08-857-076-102

Query Match 9.8%; Score 65; DB 4; Length 501;
Best Local Similarity 28.0%; Pred. No. 4.9;
Matches 33; Conservative 16; Mismatches 45; Indels 24; Gaps 7;

QY 12 GYWGCV--TEAIFETQXSLA-----ESEHXLTIGCCDADVPQPGDSRLPAVQEM 64
DB 329 GVT--GPIHTYSSLSFPAEGPLSAGCGCSSQALEAL--LTSIDTPPPADVLMTQV--- 382

QY 65 GAQEPVLDSPAIKHQLLTGDTQGRYRCRSGLSGTGMXQLSKLLETGPKVLAQSLAL 122
DB 383 ---DPLISQAPLT---LLLGCLPS---SSKLATGVCGLCPKPLFARGPSLVPITLSM 429

RESULT 10
US-08-878-989-1
; Sequence 1, Application US/08878989
; Patent No. 5885803
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl G.
; APPLICANT: Lal, Preeti
; APPLICANT: Goli, Survi K.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,989
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0321 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 685 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HUVE0801

;; CLONE: 39043
US-08-878-989-1

Query Match 9.8%; Score 65; DB 2; Length 685;
Best Local Similarity 26.7%; Pred. No. 7.5;
Matches 27; Conservative 14; Mismatches 32; Indels 28; Gaps 6;

QY 40 KTLGQ-CDAD-----VPGPGDSRLPAVQEMGAQEPVLDSPAIKHQF----- 81
DB 20 QALGKCGADSKKKRPPOPEESOPPOSO---AQP-----PAAPNNHHNHSHPETISR 71

QY 82 LTGDTQGRYRCRSGLS--GTGMXQLSKLLETGPKVLAQSL 120
DB 72 IIVDPPTTKRYCRGKVLGKGFAKCYEMTDLTNNKYAAKI 112

RESULT 11
US-09-136-282-2
; Sequence 2, Application US/09136282
; Patent No. 6063609
; GENERAL INFORMATION:
; APPLICANT: ANDERSON, KAREN
; APPLICANT: JACKSON, JEFFREY
; APPLICANT: HANSBURY, MICHAEL
; APPLICANT: NERURKAR, SANDHYA
; APPLICANT: KOSHKAR, AMY
; APPLICANT: BOUZYS, MARK
; TITLE OF INVENTION: HUMAN SERUM INDUCIBLE KINASE (SNK)
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestlia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,282
; FILING DATE: 20-AUG-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/056,112
; FILING DATE: 20-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestlia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 685 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-136-282-2

Query Match 9.8%; Score 65; DB 3; Length 685;
Best Local Similarity 26.7%; Pred. No. 7.5;
Matches 27; Conservative 14; Mismatches 32; Indels 28; Gaps 6;

QY 40 KTLGQ-CDAD-----VPGPGDSRLPAVQEMGAQEPVLDSPAIKHQF----- 81

Db 20 QALGKCGADSKKKRRPPEESOPQSO---AQP-----PAPINININSHSGPEISR 71
QY 82 LITGDPGRYRCRSGL--STGWXOLSKLLETGPKYLACSL 120
Db 72 IIVDPPTGKRYCRGKYLKGKGFACYEEMDTLTNNKYAAKI 112

RESULT 12
US-09-272-796-1
Sequence 1, Application US/09272796
Patent No. 6207148
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl G.
APPLICANT: Lal, Preeti
APPLICANT: Goll, Surya K.
APPLICANT: Shah, Puri
TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
TITLE OF INVENTION: KINASES
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/272,796
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/878,989
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0321 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 685 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HUVENOB01
CLONE: 39043
US-09-272-796-1

Query Match 9.8%; Score 65; DB 4; Length 685;
Best local Similarity 26.7%; Pred. No. 7.5;
Matches 27; Conservative 14; Mismatches 32; Indels 28; Gaps 6;

QY 40 KTLGQ-CDAD----VGRPGDSRLPAVDGMAQDERVNLDSPAIKQF----- 81
Db 20 QALGKCGADSKKKRRPPEESOPQSO---AQP-----PAPINININSHSGPEISR 71
QY 82 LITGDPGRYRCRSGL--STGWXOLSKLLETGPKYLACSL 120
Db 72 IIVDPPTGKRYCRGKYLKGKGFACYEEMDTLTNNKYAAKI 112

RESULT 13
US-09-505-744-2
Sequence 2, Application US/09505744
Patent No. 6245544
GENERAL INFORMATION:
APPLICANT: Karen M. Anderson
APPLICANT: Mark M. Bouzyk
APPLICANT: Michael J. Hansbury
APPLICANT: Jeffrey R. Jackson
APPLICANT: Sandhya S. Nerurkar
APPLICANT: Amy K. Roshak
TITLE OF INVENTION: HUMAN SERUM INDUCIBLE KINASE (SNK)
FILE REFERENCE: GH-70231-D1
CURRENT APPLICATION NUMBER: US/09/505,744
CURRENT FILING DATE: 2000-02-16
EARLIER APPLICATION NUMBER: 09/136,282
EARLIER FILING DATE: 1998-08-20
EARLIER APPLICATION NUMBER: 60/056,112
EARLIER FILING DATE: 1997-08-20
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 685
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-505-744-2

Query Match 9.8%; Score 65; DB 4; Length 685;
Best local Similarity 26.7%; Pred. No. 7.5;
Matches 27; Conservative 14; Mismatches 32; Indels 28; Gaps 6;

QY 40 KTLGQ-CDAD----VGRPGDSRLPAVDGMAQDERVNLDSPAIKQF----- 81
Db 20 QALGKCGADSKKKRRPPEESOPQSO---AQP-----PAPINININSHSGPEISR 71
QY 82 LITGDPGRYRCRSGL--STGWXOLSKLLETGPKYLACSL 120
Db 72 IIVDPPTGKRYCRGKYLKGKGFACYEEMDTLTNNKYAAKI 112

RESULT 14
US-09-345-468-9
Sequence 9, Application US/09345468
Patent No. 6245527
GENERAL INFORMATION:
APPLICANT: Busfield, S.
APPLICANT: Villedar, J.
APPLICANT: Jandrot-Perrus, M.
APPLICANT: Valinbocker, W.
TITLE OF INVENTION: GLYCOPROTEIN VI AND USES THEREOF
FILE REFERENCE: 7853-147
CURRENT APPLICATION NUMBER: US/09/345,468
CURRENT FILING DATE: 1999-06-30
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 9
LENGTH: 249
TYPE: PRT
ORGANISM: Homo sapiens
US-09-345-468-9

Query Match 9.7%; Score 64; DB 4; Length 249;
Best local Similarity 33.3%; Pred. No. 2.5;
Matches 23; Conservative 4; Mismatches 34; Indels 8; Gaps 2;

QY 51 GPRGDS--RLPAVDGMAQDERVNLDSPAIKQFLLTGDPGRYRCRSGLSTGWXOLSKL 108
Db 30 GPRGVDLYRLKLSRRYDQAVLFIAPMKRSL-----AGRYRCGYONGSLWLSLDL 83
QY 109 ELTGPKVLA 117


```

; TITLE OF INVENTION: Isolated Mammalian Monocyte Cell Genes;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,950
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,279
; FILING DATE: 21-MARCH-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,181
; FILING DATE: 16-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,252
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0670K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1204
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-985-950-20

Query Match          9.6%; Score 63.5; DB 4; Length 431;
Best Local Similarity 25.6%; Pred. No. 6.2;
Matches 33; Conservative 18; Mismatches 57; Indels 21; Gaps 5;

QY 1 MSLVFLLMGYTGWPTVEALIFETQXSLAE-----SEHLKTLGQCDADVPGP 52
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 18 MPTIVVLICLGLSPRTVHQGTIPKPTLMAEPDSVITGSPVTLSCGSLAEQ---- 73

QY 53 PGDSRLPAVQOE---WGAOEPVHLDSPAIKHQFLTGDGTGGRYRCRGLSTGKXLSK--L 107
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 74 --EYRLYREKKSASWITRIRPELVKNGQFHPSITWEHTGRIGCOYYSRARNSELSDPV 131

QY 108 LETG--PK 114
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 132 LVMTGAVPK 140

RESULT 19
US-08-985-950-16
; Sequence 16, Application US/08985950
; Patent No. 6140076
; GENERAL INFORMATION:
; APPLICANT: Adema, Gosse Jan
; TITLE OF INVENTION: Isolated Mammalian Monocyte Cell Genes;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
```

```

; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,950
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,279
; FILING DATE: 21-MARCH-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,181
; FILING DATE: 16-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,252
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0670K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1204
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 615 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-985-950-16

Query Match          9.6%; Score 63.5; DB 4; Length 615;
Best Local Similarity 25.6%; Pred. No. 10;
Matches 33; Conservative 18; Mismatches 57; Indels 21; Gaps 5;

QY 1 MSLVFLLMGYTGWPTVEALIFETQXSLAE-----SEHLKTLGQCDADVPGP 52
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 18 MPTIVVLICLGLSPRTVHQGTIPKPTLMAEPDSVITGSPVTLSCGSLAEQ---- 73

QY 53 PGDSRLPAVQOE---WGAOEPVHLDSPAIKHQFLTGDGTGGRYRCRGLSTGKXLSK--L 107
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 74 --EYRLYREKKSASWITRIRPELVKNGQFHPSITWEHTGRIGCOYYSRARNSELSDPV 131

QY 108 LETG--PK 114
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 132 LVMTGAVPK 140

RESULT 20
US-08-985-950-18
; Sequence 18, Application US/08985950
; Patent No. 6140076
; GENERAL INFORMATION:
; APPLICANT: Adema, Gosse Jan
; TITLE OF INVENTION: Isolated Mammalian Monocyte Cell Genes;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,950
FILING DATE: 05-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/041,279
FILING DATE: 21-MARCH-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/033,181
FILING DATE: 16-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/032,252
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: CHING, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0670K
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)852-9196
TELEFAX: (650)496-1204
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 615 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-985-950-18
```

```

Query Match          9.6%; Score 63.5; DB 4; Length 615;
Best Local Similarity 25.6%; Pred. No. 10;
Matches 33; Conservative 18; Mismatches 57; Indels 21; Gaps 5;

QY 1 MSMLVYVLLMGVYMGPRVDEALFYEQXSLAAE-----SEHXLKTLGQCDADVPGP 52
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 18 MPRIVTVLCLGLSLGPRTHVQGTIPKPTLMAEPDSVITQGSPTVLSCGSLDAO----- 73
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

QY 53 PGDSRLPAVOE--WGAQEPVHLSDPAIKHQFLTGTQGRYRCRSGLSGKWLQSLK--L 107
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 74 --EYRIYREKSKSWITRIRPELVKNGQFHIPSITWEHNGRGICQYSRRAWSELSDPV 131
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

QY 108 LELTG--PK 114
   |  |  |  |  |
DB 132 LVMTGAYPK 140
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 21
US-09-071-224-19
Sequence 19, Application US/09071224
Patent No. 6271343
GENERAL INFORMATION:
APPLICANT: Lok, Si
APPLICANT: Presnell, Scott R.
APPLICANT: Jelmeberg, Anna C.
APPLICANT: Gilbert, Teresa
APPLICANT: Foster, Donald C.
APPLICANT: Adams, Robyn L.
APPLICANT: Lehner, Joyce M.
TITLE OF INVENTION: MAMMALIAN ZCYTORS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics
STREET: 1201 Eastlake Ave East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
```

```

APPLICATION NUMBER: US/09/071,224
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lunn, Paul G
REGISTRATION NUMBER: 32,743
REFERENCE/DOCKET NUMBER: 96-22
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6627
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 385 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-224-19
```

```

Query Match          9.5%; Score 62.5; DB 4; Length 385;
Best Local Similarity 26.7%; Pred. No. 7.1;
Matches 32; Conservative 10; Mismatches 57; Indels 21; Gaps 5;

QY 14 TWGPVT-----EAFIFYEQXSLAAESENHL-----KTIAGQCDADVPGPDSRLPAV 61
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 152 TVGFHCHIRPKDLAF--TPYEIVTEATNRLGSARSDVLTLDVLDVYTTDPPDVHYSRV 209
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

QY 62 QENGAQEPVHLSDPAIKHQFLTGTQGRYRCRSGLSGKWLQSLKLELTGPVYLACSLA 121
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 210 GGLEDDLSVRWSPPAKDLKFLFOAKYQIRYVED--SYDMKKVDVDSNOT-----SCRLA 262
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 22
US-09-071-224-28
Sequence 28, Application US/09071224
Patent No. 6271343
GENERAL INFORMATION:
APPLICANT: Lok, Si
APPLICANT: Presnell, Scott R.
APPLICANT: Jelmeberg, Anna C.
APPLICANT: Gilbert, Teresa
APPLICANT: Foster, Donald C.
APPLICANT: Adams, Robyn L.
APPLICANT: Lehner, Joyce M.
TITLE OF INVENTION: MAMMALIAN ZCYTORS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics
STREET: 1201 Eastlake Ave East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,224
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lunn, Paul G
REGISTRATION NUMBER: 32,743
```



```

1      MEDION TYPE: Diskette
2      COMPUTER: IBM compatible
3      OPERATING SYSTEM: DOS
4      SOFTWARE: FASTSEQ for Windows Version 2.0
5      CURRENT APPLICATION DATA:
6      APPLICATION NUMBER: US/09/071,224
7      FILING DATE:
8      CLASSIFICATION:
9      PRIOR APPLICATION DATA:
10     APPLICATION NUMBER:
11     FILING DATE:
12     ATTORNEY/AGENT INFORMATION:
13     NAME: Lunn, Paul G
14     REGISTRATION NUMBER: 32,743
15     REFERENCE/DOCKET NUMBER: 96-22
16     TELECOMMUNICATION INFORMATION:
17     TELEPHONE: 206-442-6627
18     TELEFAX: 206-442-6678
19     TELEX:
20     INFORMATION FOR SEQ ID NO: 22:
21     SEQUENCE CHARACTERISTICS:
22     LENGTH: 389 amino acids
23     TYPE: amino acid
24     STRANDEDNESS: single
25     TOPOLOGY: linear
26     MOLECULE TYPE: protein
27     US-09-071-224-22

```

Query Match	9.3%	Score 61.5;	DB 4;	Length 389;
Best Local Similarity	26.7%;	Pred. NO. 9.8;		
Matches 32;	Conservative 10;	Mismatches 57;	Indels 21;	Gaps 5

0y	14	TMGPTV-----EAMIFPTQASLWASESHXJ-----KTLTGQCDADVPQPGGSTRLP	PAV	61
Db	152	TYGHRSCHIRPDALF--TPREIVNEAENRGLSARSDVLTILDLIDVYTTDPDPVHWHSRV		209
0y	62	QEWKAGPEVHLDPARKIQHQLLTDTGQGRVCRSGSLSTGMXQSLLLELTGPKVLA	CLSA	121
2b	210	CGLEHQLSVRWVSPALKDPLFQAKYQIARYVED--SYDMKVVVDVDSNQ-----SCRLA		262

RESULT 29
US-09-071-224-24
; Sequence 24, Application US/09071224

APPLICANT: Lok, Si
APPLICANT: Pirenell, Scott R.
APPLICANT: Jelmberg, Anna C.
APPLICANT: Gilbert, Teresa
APPLICANT: Foster, Donald C.
APPLICANT: Adams, Robyn L.
APPLICANT: Lehner, Joyce M.
TITLE OF INVENTION: MAMMALIAN ZCYTORS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics
STREET: 1201 Eastlake Ave East
CITY: Seattle

```

1
2
3      COMPUTER READABLE FORM:
4      MEDIUM TYPE: Diskette
5
6      COMPUTER: IBM Compatible
7
8      OPERATING SYSTEM: DOS
9
10     SOFTWARE: FASTSEQ for Windows Version 2.0
11
12     CURRENT APPLICATION DATA:
13
14     APPLICATION NUMBER: US/09/071,224
15
16     FILING DATE:
17
18     CLASSIFICATION:
19
20     PRIOR APPLICATION DATA:
21
22

```

```

1 APPLICATION NUMBER:
2
3 FILING DATE:
4
5 ATTORNEY/AGENT INFORMATION:
6
7 NAME: Lunn, Paul G
8
9 REGISTRATION NUMBER: 32,743
10
11 REFERENCE/DOCKET NUMBER: 96-22
12
13 TELECOMMUNICATION INFORMATION
14
15 TELEPHONE: 206-442-6627
16
17 TELEFAX: 206-442-6678
18
19 TELEX:
20
21 INFORMATION FOR SEQ ID NO: 24:
22
23 SEQUENCE CHARACTERISTICS:
24
25 LENGTH: 389 amino acids
26
27 TYPE: amino acid
28
29 STRANDEDNESS: single
30
31 TOPOLOGY: linear
32
33 MOLECULE TYPE: protein
34
35 OS-09-071-224-24

```

Query Match	9.38;	Score	61.5;	DB	4;	Length	389;
Best Local Similarity	26.78;	Pred. No.	9.8;				
Matches	32;	Conservative	10;	Mismatches	57;	Indels	21;
						Gaps	5

[illegible]

RESULT 30
US-09-071-224-25
; Sequence 25, Application US/09071224

1 GENERAL INFORMATION:
 2 APPLICANT: Lok, S1
 3 APPLICANT: Pressnell, Scott R.
 4 APPLICANT: Jelmberg, Anna C.
 5 APPLICANT: Gilbert, Teresa C.
 6 APPLICANT: Foster, Donald C.
 7 APPLICANT: Adams, Royn L.
 8 APPLICANT: Lehner, Joyce M.
 9 TITLE OF INVENTION: MAMMALIAN ZYGOTORES
 10 NUMBER OF SEQUENCES: 37
 11 CORRESPONDENCE ADDRESS:
 12

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525


```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,965
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/348,353
FILING DATE: 30-NOV-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/247,572
FILING DATE: 23-MAY-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/03725
FILING DATE: 04-MAY-1992
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/695,613
FILING DATE: 03-MAY-1991
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Jackson, David
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-097CIP1DIV2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1248 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-465-965-17

Query Match          9.3%; Score 61.5; DB 2; Length 1248;
Best Local Similarity 30.6%; Pred. No. 49;
Matches 22; Conservative 13; Mismatches 28; Indels 9; Gaps 3

QY   26 ETQXSLAESEHXLK--TLGQCDAVPGPFGDSRLPA---VOENG---AQEPVHLDSFA 76
      | ||| | : : : : : : : : : : : : : : : : : : : : : : : : :
DB   879 EANALIMAGELTVKAQNITKRALLIEAGNARLTAAVALINKIGRIACGDMDHLDAPR 938

QY   77 IKHQFLTGTDPQ 88
      |::|::|:|
DB   939 IENTAKLGEVQ 950

RESULT 36
US-08-465-966-17
Sequence 17, Application US/08465966
Patent No. 6015360
GENERAL INFORMATION:
Applicant: Tuomanen, Elaine
Title of Invention: Antibody Recognizing Endothelial Cell
Number of Sequences: 38
TITLE OF INVENTION: Ligand for Leukocyte CR3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klausner & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,966
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/348,353
FILING DATE: 30-NOV-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/247,572
FILING DATE: 23-MAY-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/03725
FILING DATE: 04-MAY-1992
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/695,613
FILING DATE: 03-MAY-1991
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Jackson, David
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-097C1P1D1V2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1248 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-465-966-17

Query Match          9.3%; Score 61.5; DB 3: Length 1248; *
Best Local Similarity 30.6%; Pred. No. 49;
Matches 22; Conservative 13; Mismatches 28; Indels 9; Gaps 3;

OY 26 EFTQSLMAESEHLK*-TLTGCCDAVPPGPGDSRLP-----VQEWG---AOEPIHLDSPA 76
      |  | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 879 ENALLMLMAGELTVAAONTTKRAALIEAGGNARLTPAAVALINKGRIRACEDMHLDAFR 938
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY 77 IKHQFLITGDTQ 88
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 939 IENTAKLSEGV 950

RESULT 37
US-07-642-734C-4
: Sequence 4, Application US/07642734C
: Patent No. 5824513
: GENERAL INFORMATION:
: APPLICANT: Katz, L
: APPLICANT: Donadio, S
: APPLICANT: McAlpine, J B
: TITLE OF INVENTION: Recombinant DNA Method for Producing
: TITLE OF INVENTION: Erythromycin Analogs
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Edward H. Gorman
: STREET: Abbott Laboratories D377/AP6D-2 One Abbott
: CITY: Abbott Park
: STATE: IL
: COUNTRY: US
: ZIP: 60064-3500
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25

```


Db 152 TWGPHSCHIPKDLALF--TPYEIWEATNRNLGARSDDLTLIDIVTTIDPPDVHVSRY 209
QY 62 QEMGAQEPHYDSPAIFKHOFLTLTGDTGGRYRCRSGSLTGMXOLSKLELTGPKVILACSLA 121
Db 210 GGLEDDLSVYRWVSPALKDFLFOAKYQIRYVED--SYDMKVVDVSNQI-----SCRLA 262

RESULT 40
US-08-449-645A-19
; Sequence 19, Application US/08449645A
; Patent No. 5981245
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: Eph-Like Receptor Protein Tyrosine
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Patent Operations/RBW
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,645A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-287
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-449-645A-19

Query Match 9.2%; Score 60.5; DB 2; Length 984;
Best Local Similarity 23.1%; Pred. No. 48;
Matches 31; Conservative 13; Mismatches 45; Indels 45; Gaps 6;

QY 17 PVEALIFVETQXSLWA-----ESPHXKLTIGQCDADY----PGP 52
Db 856 PVCCPAPLIELMKNCWAYDRARRHFQKLOAHLEOLLANPHSLRTIANFDPRVTLRLPSL 915

QY 53 PGDSRLP--AVQEMGAQEPHYDSPAIFKHOFLTLTGDTGGRYRCRSGSLTGMXOLSKLEL 110
Db 916 SGSDGIPRYTVSEW-----LESIMKRYIL-----HFHSAGLDI----MECVLEL 956

QY 111 TGPKVILACSLALDG 124
Db 957 TAEDLTQMGITLPG 970

RESULT 41
US-08-702-367A-19
; Sequence 19, Application US/08702367A
; Patent No. 5981246
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: Eph-Like Receptor Protein Tyrosine
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Patent Operations/RBW

STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/702,367A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-287
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-702-367A-19

Query Match 9.2%; Score 60.5; DB 2; Length 984;
Best Local Similarity 23.1%; Pred. No. 48;
Matches 31; Conservative 13; Mismatches 45; Indels 45; Gaps 6;

QY 17 PVEALIFVETQXSLWA-----ESPHXKLTIGQCDADY----PGP 52
Db 856 PVCCPAPLIELMKNCWAYDRARRHFQKLOAHLEOLLANPHSLRTIANFDPRVTLRLPSL 915

QY 53 PGDSRLP--AVQEMGAQEPHYDSPAIFKHOFLTLTGDTGGRYRCRSGSLTGMXOLSKLEL 110
Db 916 SGSDGIPRYTVSEW-----LESIMKRYIL-----HFHSAGLDI----MECVLEL 956

QY 111 TGPKVILACSLALDG 124
Db 957 TAEDLTQMGITLPG 970

RESULT 42
PCT-US95-04681-19
; Sequence 19, Application PCT/US9504681
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: Eph-Like Receptor Protein Tyrosine
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Patent Operations/RBW
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04681
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-287
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:

LENGTH: 984 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-04681-19

Query Match 9.2%: Score 60.5; DB 5; Length 984;
Best Local Similarity 23.1%: Pred. No. 48;
Matches 31; Conservative 13; Mismatches 45; Indels 45; Gaps 6;

QY 17 PYTEAIFETQXSLWA-----ESEHXIKTLGQCDADV-----PGP 52
DB 856 PVCCPAPLYELMKNCWADRRPHQKQAHLEQLLANHSLRTIANFDPRYTLRLPSL 915
QY 53 PCDSRLP--AVQEWGAQEPVHLDSPAIKHQFLLTGTGCRGRSGLSGTWQSLKLEL 110
DB 916 SCSDGIPRYTVEW-----LESIRMKRYTL-----HFSACLDT-----MECVLEL 956
QY 111 TSPKVLACSLALDG 124
DB 957 TAEDLTQMGITLPG 970

RESULT 43
US-08-667-939A-2
Sequence 2, Application US/08667939A
Patent No. 5998166
GENERAL INFORMATION:
APPLICANT: LUD, Shun
TITLE OF INVENTION: CD16-II VARIANTS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NETMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/667,939A
FILING DATE: 24-JUN-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/433,123
FILING DATE: 03-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: LUD 2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 254 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-667-939A-2

Query Match 9.1%: Score 60; DB 2; Length 254;
Best Local Similarity 28.3%: Pred. No. 8.5;
Matches 34; Conservative 13; Mismatches 43; Indels 30; Gaps 7;
QY 3 MLVVFLLMGVTVGPVTE---AAIFVETO-XSLMAESEHXIKTLGQCDADVPGPPGDSR 57

DB 4 LLLPTALLLLVSAGMRTEDELPKAVFLEPQWYSVLEKDSVTLKQCAYS-----PEDNS 57
QY 58 LPRAVQEWGAQEPVHLDSPAIKHQFL--LTGDTGCRGRSGLSGT-----GWXL 104
DB 58 ---TQMFHKE--NLISSQASSYFIDAATVDDSGEYRCQTNISTLSDPVQLLEVQVGMILL 111

RESULT 44
US-08-759-628-5
Sequence 5, Application US/08759628
Patent No. 6225446
GENERAL INFORMATION:
APPLICANT: Altmann, Scott W.
APPLICANT: Rock, Fernando L.
APPLICANT: Bazan, J. Fernando
APPLICANT: Kastelein, Robert A.
TITLE OF INVENTION: MUTATIONAL VARIANTS OF MAMMALIAN PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNA Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/759,628
FILING DATE: 05-DEC-1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/008,574
FILING DATE: 06-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX05520
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-852-9196
TELEFAX: 415-496-1200
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-759-628-5

Query Match 9.1%: Score 60; DB 4; Length 421;
Best Local Similarity 23.7%: Pred. No. 17;
Matches 40; Conservative 14; Mismatches 59; Indels 56; Gaps 9;

QY 4 LVVVFLLMGVTVGPVTEAIFVETOXSLMAESEHXIKTLGQCDADV-----GPP----- 53
DB 3 LLLPLLMG---GSLQEKPY-YELQVOKSVTVQEGLCVLVPCFSYPMRWSWSSPLYVY 58
QY 54 --GDSRLPAVQEWGA-QEPVHLDSPAIKHQFLLTGTGCR-----YR 92
DB 59 WFRDGEIPYVAEVVAANNPDRRVKPTQGRFRLGDVOKKNCSLSTGDARMEBDTGSYFPR 118
QY 93 CRGSLSTGWX-QLSKL--LELTG-----PKVLACSL 120
DB 119 VENGRIYKSYQONKLNLEVTALEKPDILHLSGPLESGLMRPRLSCSL 167

RESULT 45

US-07-683-957B-3
; Sequence 3, Application US/07683957B
; Patent No. 5310880
; GENERAL INFORMATION:
; APPLICANT: Donahoe, Patricia K.
; APPLICANT: Ragin, Richard C.
; APPLICANT: MacLaughlin, David T.
; TITLE OF INVENTION: Purification of M llerian Inhibiting
; TITLE OF INVENTION: Substance
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/683,957B
; FILING DATE: 19910412
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldstein, Jorge A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 0609,3060000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 466-0800
; TELEFAX: (202) 833-8716
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-683-957B-3

Query Match 9.1%; Score 60; DB 1; Length 553;
Best Local Similarity 27.8%; Pred. No. 25;
Matches 32; Conservative 9; Mismatches 52; Indels 22; Gaps 6;
QY 20 EAALFYETQXSLMASEHXIKTLGQCDDADVPGPQSDRLPAVO-----EWGADE--PV 70
| | | | | : | | | | | : | | | | : | | :
DB 77 EQAFLEAVQESRMGPQD--LATFGVCSFD----SQTLPALQRLGAWLGGTGEQQLAVL 129
| | | | | : | | | | | : | | | | : | | :
QY 71 HDSIPAIIKHQFLITGDTQGRYRCRSGSLSTGWXO-LSKLLLELTGPKVLAACSLALOG 124
| | | | | : | | | | | : | | | | : | | :
DB 130 HLAETVMEPOLLLKFOEP-----PRGASRWQALIVLPGPQPVTVTGAGILOG 179
| | | | | : | | | | | : | | | | : | | :

Search completed: January 7, 2002, 16:49:55
Job time: 282 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2002, 16:44:08 ; Search time 31.99 Seconds

(without alignments)
291.755 Million cell updates/sec

Title: US-09-471-276-831

Perfect score: 661

Sequence: 1 MSMLVFLLLMGVTWGPVTE.....LLELTGPKVLACSLALDGAS 126

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

A_Geneseq_1101.*
1: /SID52/gcgcdata/geneseq/geneseq/AA1980.DAT.*
2: /SID52/gcgcdata/geneseq/geneseq/AA1981.DAT.*
3: /SID52/gcgcdata/geneseq/geneseq/AA1982.DAT.*
4: /SID52/gcgcdata/geneseq/geneseq/AA1983.DAT.*
5: /SID52/gcgcdata/geneseq/geneseq/AA1984.DAT.*
6: /SID52/gcgcdata/geneseq/geneseq/AA1985.DAT.*
7: /SID52/gcgcdata/geneseq/geneseq/AA1986.DAT.*
8: /SID52/gcgcdata/geneseq/geneseq/AA1987.DAT.*
9: /SID52/gcgcdata/geneseq/geneseq/AA1988.DAT.*
10: /SID52/gcgcdata/geneseq/geneseq/AA1989.DAT.*
11: /SID52/gcgcdata/geneseq/geneseq/AA1990.DAT.*
12: /SID52/gcgcdata/geneseq/geneseq/AA1991.DAT.*
13: /SID52/gcgcdata/geneseq/geneseq/AA1992.DAT.*
14: /SID52/gcgcdata/geneseq/geneseq/AA1993.DAT.*
15: /SID52/gcgcdata/geneseq/geneseq/AA1994.DAT.*
16: /SID52/gcgcdata/geneseq/geneseq/AA1995.DAT.*
17: /SID52/gcgcdata/geneseq/geneseq/AA1996.DAT.*
18: /SID52/gcgcdata/geneseq/geneseq/AA1997.DAT.*
19: /SID52/gcgcdata/geneseq/geneseq/AA1998.DAT.*
20: /SID52/gcgcdata/geneseq/geneseq/AA1999.DAT.*
21: /SID52/gcgcdata/geneseq/geneseq/AA2000.DAT.*
22: /SID52/gcgcdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	655	99.1	126	21	AAV64670 Human 5' EST relat
2	139.5	21.1	512	22	ABA48724 Mouse liver growth
3	95.5	14.4	299	22	AAU01393 Human MANGO 511, v
4	92.5	14.0	299	22	AAU01394 Human MANGO 511, v
5	91.5	13.8	265	21	AAU04182 Leukocyte immunogl
6	91.5	13.8	299	21	AAU04180 Leukocyte immunogl
7	91.5	13.8	299	22	AAU01330 Human MANGO 511 am
8	91.5	13.8	299	22	AAU01391 Human MANGO 511, v
9	91.5	13.8	299	22	AAU01392 Human MANGO 511, v
10	90.5	13.7	623	19	AAW69233 FCR-III protein se
11	90.5	13.7	631	19	AAW82552 Human LIR-pbm17 pr

12	90.5	13.7	631	21	AAU04177 Leukocyte immunogl
13	87.5	13.2	631	22	AAU61263 Human monocyte inh
14	85.5	12.9	472	19	AAW69234 FCR-IV protein seq
15	84.5	12.8	448	19	AAW82551 Human LIR-pbm2 pro
16	84.5	12.8	448	19	AAW53466 Human gp49 HM18 po
17	84.5	12.8	448	21	AAU04176 Leukocyte immunogl
18	81.5	12.3	590	19	AAW82553 Human LIR-pbmnew p
19	81.5	12.3	590	21	AAU04178 Leukocyte immunogl
20	78.5	11.9	652	19	AAW82545 Human LIR-18A3 pro
21	78.5	11.9	652	21	AAU04170 Leukocyte immunogl
22	75.5	11.4	244	21	AAV57323 Nostoc NspI restr
23	75.5	11.4	635	20	AAV24321 Mouse dephosphoryl
24	75	11.3	458	22	AAU29500 Human protein sequ
25	75	11.3	647	22	AAU64131 Leucine multi-repe
26	74.5	11.3	663	20	AAV24322 Mouse dephosphoryl
27	74.5	11.3	680	20	AAV24320 Mouse dephosphoryl
28	74	11.2	289	19	AAW82548 Human LIR-pbm36-2
29	74	11.2	289	21	AAU04173 Leukocyte immunogl
30	74	11.2	650	19	AAW82544 Human LIR-P3G2 pro
31	74	11.2	650	21	AAU04169 Leukocyte immunogl
32	74	11.2	651	19	AAW62782 Protein encoded by
33	73	11.0	122	19	AAW41238 Alpha-v-beta-5 ant
34	73	11.0	193	19	AAW41090 Chicken matrix met
35	73	11.0	193	19	AAW41235 Alpha-v-beta-5 ant
36	73	11.0	228	19	AAW41089 Chicken matrix met
37	73	11.0	228	19	AAW41234 Alpha-v-beta-5 ant
38	73	11.0	663	19	AAW41111 Chicken matrix met
39	73	11.0	663	19	AAW41227 Chicken matrix met
40	72.5	11.0	244	21	AAU29926 Nostoc sp restrict
41	72.5	11.0	253	21	AAU04183 Leukocyte immunogl
42	72.5	11.0	287	21	AAU04181 Leukocyte immunogl
43	72	10.9	542	20	AAV15104 Chimeric protein I
44	71.5	10.8	90	22	AAW25575 Human protein sequ
45	71	10.7	324	12	AAU12427 Hybrid Fc(gamma)RI
46	70	10.6	192	21	AAU69300 HIV-1 non-subtype
47	69.5	10.5	341	22	AAU65833 Murine mature INTR
48	69.5	10.5	370	22	AAU65832 Human INTERCEPT
49	69.5	10.5	370	22	AAU65904 Human secreted pro
50	69.5	10.5	370	22	AAU65905 Human secreted pro
51	69.5	10.5	841	20	AAV24318 Mouse dephosphoryl
52	69	10.4	339	22	AAV27290 Human platelet mem
53	69	10.4	489	19	AAW82549 Human LIR-pbm36-4
54	69	10.4	489	21	AAU04174 Leukocyte immunogl
55	68.5	10.4	244	21	AAU42513 Arabidopsis thaila
56	68.5	10.4	321	22	AAU40551 Human polypeptide
57	68.5	10.4	325	21	AAV95024 Human secreted pro
58	68.5	10.4	389	21	AAV76303 Human viral recept
59	68.5	10.4	390	20	AAV27096 Human acid sequen
60	68.5	10.4	390	20	AAV13351 EGF-like homologue
61	68.5	10.4	390	20	AAU05286 Human secreted pro
62	68.5	10.4	390	21	AAU88574 Human secreted pro
63	68.5	10.4	390	21	AAU94999 Human secreted pro
64	68.5	10.4	390	22	AAU12340 Human PRO246 polyp
65	68.5	10.4	390	22	AAU06610 Human protein havi
66	68.5	10.4	390	22	AAU90818 Human shear stress
67	68.5	10.4	390	22	AAU88358 Human membrane or
68	68.5	10.4	390	22	AAU88359 PRO246. Homo sapi
69	68.5	10.4	390	22	AAU88360 Amino acid sequen
70	68.5	10.4	390	22	AAU88361 Human PRO246 prote
71	68.5	10.4	390	22	AAU88362 Human anglogenesis
72	68	10.3	339	22	AAU61274 Human TANGO 268-re
73	68	10.3	775	22	AAU23632 Human EST encoded
74	67.5	10.2	119	21	AAU54198 Human pancreatic c
75	67.5	10.2	141	22	AAU39267 Human polypeptide
76	67.5	10.2	145	22	AAU41053 Human polypeptide
77	67.5	10.2	217	22	AAU65871 Human INTERCEPT
78	67.5	10.2	220	22	AAU65872 Human INTERCEPT
79	67.5	10.2	246	22	AAU65835 Murine INTERCEPT
80	67.5	10.2	370	22	AAU65906 Human secreted pro
81	67	10.1	16	22	Schizophrenia-asso
82	67	10.1	16	22	AAU15679 Breast-cancer asso
83	67	10.1	16	22	AAU87213 Breast-cancer asso
84	67	10.1	16	22	AAU87241 Breast-cancer asso

```
85 67 10.1 254 21 AAY58671 Porcine reproduct
86 66.5 10.1 219 22 AAM25945 Human protein sequ
87 66.5 10.1 503 20 AAY15103 Modified interfero
88 66.5 10.1 506 21 AAY94986 Human secreted pro
89 66.5 10.1 1081 20 AAY24319 Mouse dephosphoryl
90 66 10.0 342 22 AAB48740 Mouse liver growth
91 66 10.0 415 15 AAR54946 Cuphea hookeriana
92 66 10.0 415 16 AAR74145 Cuphea class II th
93 66 10.0 415 19 AAM44335 Class II Cuphea ho
94 65.5 9.9 370 22 AAB65907 Human secreted pro
95 65 9.8 12 22 AAB15668 Schizophrenta-asso
96 65 9.8 12 22 AAB15680 Schizophrenta-asso
97 65 9.8 12 22 AAB87217 Breast-cancer asso
98 65 9.8 12 22 AAB87245 Breast-cancer asso
99 65 9.8 131 22 AAM43502 Human polypeptide
100 65 9.8 501 21 AAY96447 Forkhead transcrip
```

ALIGNMENTS

RESULT 1

ID AAY64670 standard; Protein; 126 AA.

XX AAY64670:

DT 01-FEB-2000 (first entry)

DE Human 5' EST related polypeptide SEQ ID NO:831.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
XX regulation; identification.

OS Homo sapiens.

PN WO953051-A2.

PD 21-OCT-1999.

PF 09-APR-1999; 99WO-1B00712.

PR 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

XX (GEST) GENSET.

PA Dumas Milne Edwards J, Duclert A, Giordano J;

PI MPI: 2000-038446/03.

DR N-PSDB: AAZ42284.

PT Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures

PS Claim 3; Page 603; 837pp; English.

XX AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AAY5438 represent the EST-related proteins corresponding to AAZ42265 to
CC AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the

CC Insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.

SQ Sequence 126 AA:

Query Match 99.1%; Score 655; DB 21; Length 126;

Best local similarity 100.0%; Pred. No. 2,1e-70; Mismatches 0; Indels 0; Gaps 0;

Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSMLVVFLLMGVTGCPVPEAIFEFYFQXSLMAESEHXLKTLGOCADVDGPPGDSRLPA 60

Db 1 msmlvlflllmgvtgcpvpeaalfeyqqsllwesehxlkllygcadadvppgdsrlpa 60

OY 61 VQEMGAQEPVHLDPSPAIKHQFLTLTGDTGCRRCRSGLSGTGMXLSKILTLGPKVLACST 120

Db 61 vqewgaqepvhlspalkhqlflltgdtggrtcrsglsqwxlsklltlgpkvlacsl 120

OY 121 ALDGAS 126

Db 121 aldgas 126

RESULT 2

ID AAB48724 standard; Protein; 512 AA.

XX AAB48724:

DT 09-MAR-2001 (first entry)

DE Mouse liver growth hormone-induced clone 5 ORF #1, SEQ ID NO:8.

XX Mouse; growth hormone; GH regulatable gene; liver pathology; hypertrophy;
KW hepatocellular lesion; hyperplasia; altered expression level; clone 5;
KW diagnostic marker; gigantism; acromegaly; diabetes; hepatotropic;
KW transgenic animal; drug screening; drug discovery; murine;
KW cytoplasmic protein; open reading frame; ORF.

OS Mus sp.

PN WO200066787-A2.

PD 09-NOV-2000.

PF 05-MAY-2000; 2000WO-US12366.

PR 05-MAY-1999; 99US-0132663.

XX (UYOH-) UNIV OHIO.

PA Kopchick JJ, Tjong J;

PI MPI: 2001-007239/01.

DR N-PSDB: AAC87257.

PT Diagnosing abnormal levels of growth hormone activity in liver
PT comprising assaying growth transcriptional activity and protein
PT expression level of hormone-regulatable liver genes, as diagnostic
PT markers of liver pathology

PS Claim 2; Page 53; 65pp; English.

XX The invention relates to a method of diagnosing abnormal levels of
CC growth hormone (GH) activity in the liver, or predicting a change in the
CC condition of the liver in response to abnormal GH activity. The method
CC involves correlating the level of expression of certain specific genes
CC with the level of GH activity in the liver, or with an expected change
CC in the condition of the liver as the result of GH activity. Excessive GH

XX	WO200121631-A2.
PN	
PD	29-MAR-2001.
XX	
XX	20-SEP-2000; 2000WO-US25982.
PF	
PR	20-SEP-1999; 99US-039972.
XX	
PA	(MILL.) MILLENNIUM PHARM INC.
XX	
PI	Kirst SJ, Sharp JD, Fraser CC, Barnes T, Kingsbury G;
DR	WPI: 2001-211461/21.
DR	N-PSDB: AAS02104.
XX	
PT	New nucleic acid encoding INTERCEPT 307, MANGO 511, TANGO 351, TANGO
PT	361, TANGO 499 or TANGO 509 secreted or transmembrane protein, useful
PT	for the diagnosis and treatment of arthritis, psoriasis and Parkinson's
PT	disease -
XX	
PS	Disclosure; Page 299-300; 362pp; English.
XX	
CC	The sequence represents the amino acid sequence of human MANGO 511
CC	variant #3 transmembrane protein. The nucleic acid and polypeptide
CC	sequences are useful for the diagnosis, prognosis and treatment of
CC	immunological disorders (e.g. arthritis, graft rejection and acquired
CC	immunodeficiency syndrome), inflammatory disorders (e.g. psoriasis and
CC	asthma), renal disorders, embryonic disorders, brain-related disorders
CC	(e.g. cerebral oedema), cerebrovascular diseases (e.g. ischaemia),
CC	tumour, prostate-related disorders, pituitary-related disorders (e.g.
CC	Cushing's disease) and neurodegenerative diseases (e.g. Parkinson's

SQ Sequence 239 AA:

```

QY      1 MSLVVELLLMGCVTMGPTVEAAIFETQXSLMAE-----SEHXLKLTKGCGAD---- 48
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      19 vspalmlilcglclprthvgaguskatlaepgsvsrgsvtlrcqglleaeyrl 78
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      49 -VPGPGCDRLAVNDEMGMQEPVHLDSPAIKQIFLLTGTC---ERYRNRSGLSGMXL 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      79 vkegsp-----epwdtgmpl---epankarfslpsmtlehagryfcyyyspagwsep 127
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      128 sdpiel 133
XX
XX      RESULT      4
XX      AAU01394
XX      ID      AAU01394 standard; Protein; 299 AA.
XX      AC      AAU01394;
XX      DT      18-JUL-2001 (first entry)
XX      DE      Human MANGO 511, variant #4 amino acid sequence.
XX      KW      Human; MANGO 511; transmembrane protein; diagnostic; asthma;
XX      KW      immunological disorder; arthritis; graft rejection; renal disorder;
XX      KW      acquired immunodeficiency syndrome; inflammatory disorders; psoriasis;
XX      KW      AIDS; embryonic disorder; brain; cerebral oedema; ischaemia; tumour;
XX      KW      prostate; cerebrovascular disease; pituitary; Cushing's disease;
XX      KW      neurodegenerative disease; Parkinson's disease.
XX      OS      Homo sapiens.
XX

```

PN WO200121631-A2.
XX
PD 29-MAR-2001.
XX
PF 20-SEP-2000; 2000WO-US25982.
XX
PR 20-SEP-1999; 99US-0399723.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Kirst SJ, Sharp JD, Fraser CC, Barnes T, Kingsbury G;
XX WPI: 2001-211461/21.
DR N-PSDB; AAS02105.
XX
PT New nucleic acid encoding INTERCEPT 307, MANGO 511, TANGO 351, TANGO
PT 361, TANGO 499 or TANGO 509 secreted or transmembrane protein, useful
PT for the diagnosis and treatment of arthritis, psoriasis and Parkinson's
PT disease -
XX
PS Disclosure: Page 301-302; 362pp; English.
XX
CC The sequence represents the amino acid sequence of human MANGO 511
CC variant #4 transmembrane protein. The nucleic acid and polypeptide
CC sequences are useful for the diagnosis, prognosis and treatment of
CC immunological disorders (e.g. arthritis, graft rejection and acquired
CC immunodeficiency syndrome), inflammatory disorders (e.g. psoriasis and
CC asthma), renal disorders, embryonic disorders, brain-related disorders
CC (e.g. cerebral oedema), cerebrovascular diseases (e.g. ischaemia),
CC tumours, prostate-related disorders, pituitary-related disorders (e.g.
CC Cushing's disease) and neurodegenerative diseases (e.g. Parkinson's
CC disease).
SQ Sequence 299 AA:

Query Match 14.0%; Score 92.5; DB 22; Length 299;
Best Local Similarity 26.2%; Pred. No. 0.0088;
Matches 33; Conservative 18; Mismatches 48; Indels 27; Gaps 5;

OY 1 MSMLVFLILMGVTGPTVEAIFETQXSLMAE-----SEHLKTLGQCDAD----- 48
DB 19 VSPALMVLICIGLSIGPKTHVAGNLSKATLWEPGVSIRGNSVILRCGTLAEAGEYTL 78
OY 49 -VPGPGDSRLPAVQEWGAQEPVHLDSPAIKHOFILTGDTQ---GRYRCRSGLSGTGWXOL 104
DB 79 VKEGSP-----EPWDTQNP1---EPKHKARFISIPSMTEHAGTYRCYYSPGWSESP 127
OY 105 SKLLEL 110
DB 128 SDPLEL 133

RESULT 5
AAB04182
ID AAB04182 standard; Protein: 265 AA.
XX
AC AAB04182;
XX
DT 11-APR-2001 (first entry)
XX
DE Leukocyte immunoglobulin like receptor LIR-9s1.
XX
KM Leukocyte immunoglobulin like receptor; LIR; gene therapy;
KM autoimmunity; autoimmune disorders; immune system; human.
OS Homo sapiens.
XX
PN WO200068383-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US13228.

XX
PR 12-MAY-1999; 99US-0310463.
XX
PA (IMNV) IMMUNEX CORP.
XX
PI Cosman DJ, Anderson DM, Borges L;
XX WPI: 2000-687645/67.
DR N-PSDB; AAS54610.
XX
PT Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
PT treating autoimmune diseases and disease states with suppressed immune
PT function
XX
PS Claim 1: Page 111; 117pp; English.
XX
CC Host cells transformed with a vector which are capable of
CC expressing a leukocyte immunoglobulin like receptor (LIR), can be
CC used to produce the LIR polypeptide. LIR coding sequences may be
CC used in the gene therapy of disorders mediated directly or
CC indirectly by defective or insufficient amounts of any of the LIR
CC polypeptides. The LIR polypeptides can be used to treat autoimmune
CC diseases and disease states with suppressed immune function.
SQ Sequence 265 AA:

Query Match 13.8%; Score 91.5; DB 21; Length 265;
Best Local Similarity 26.2%; Pred. No. 0.0099;
Matches 33; Conservative 18; Mismatches 48; Indels 27; Gaps 5;

OY 1 MSMLVFLILMGVTGPTVEAIFETQXSLMAE-----SEHLKTLGQCDAD----- 48
DB 19 VSPALMVLICIGLSIGPKTHVAGNLSKATLWEPGVSIRGNSVILRCGTLAEAGEYTL 78
OY 49 -VPGPGDSRLPAVQEWGAQEPVHLDSPAIKHOFILTGDTQ---GRYRCRSGLSGTGWXOL 104
DB 79 VKEGSP-----EPWDTQNP1---EPKHKARFISIPSMTEHAGTYRCYYSPGWSESP 127
OY 105 SKLLEL 110
DB 128 SDPLEL 133

RESULT 6
AAB04180
ID AAB04180 standard; Protein: 299 AA.
XX
AC AAB04180;
XX
DT 11-APR-2001 (first entry)
XX
DE Leukocyte immunoglobulin like receptor LIR-9m1.
XX
KM Leukocyte immunoglobulin like receptor; LIR; gene therapy;
KM autoimmunity; autoimmune disorders; immune system; human.
OS Homo sapiens.
XX
PN WO200068383-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US13228.
XX
PR 12-MAY-1999; 99US-0310463.
XX
PA (IMNV) IMMUNEX CORP.
XX
PI Cosman DJ, Anderson DM, Borges L;
XX WPI: 2000-687645/67.
DR N-PSDB; AAS54608.

FT /note= "N-myristylation site"

QY	103	54162	110
QY	103	54162	110

Db	128	sdple1	133
----	-----	--------	-----

```
RESULT 8
AAU01391
ID AAU01391 standard; Protein; 299 AA.
XX
AC AAU01391;
XX
DT 18-JUL-2001 (first entry)
XX
DE Human MANCO 511, variant #1 amino acid sequence.
XX
KW Human; MANCO 511; transmembrane protein; diagnostic; asthma;
KW immunological disorder; arthritis; graft rejection; renal disorder;
KW acquired immunodeficiency syndrome; inflammatory disorders; psoriasis;
KW AIDS; embryonic disorder; brain; cerebral oedema; ischaemia; tumour;
KW prostate; cerebrovascular disease; pituitary; Cushing's disease;
KW neurodegenerative disease; Parkinson's disease.
XX
OS Homo sapiens.
XX
PN WO200121631-A2.
XX
PD 29-MAR-2001.
XX
PF 20-SEP-2000; 2000WO-US25982.
XX
PR 20-SEP-1999; 99US-0399723.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Kirst SJ, Sharp JD, Fraser CC, Barnes T, Kingsbury G;
XX
DR MPI: 2001-211461/21.
XX
DR N-PSDB: AAS02102.
XX
PT New nucleic acid encoding INTERCEPT 307, MANCO 511, TANGO 351, TANGO
PT 361, TANGO 499 or TANGO 509 secreted or transmembrane protein, useful
PT for the diagnosis and treatment of arthritis, psoriasis and Parkinson's
PT disease -
XX
XX
PS Disclosure: Page 295-296; 362pp; English.
XX
XX
CC The sequence represents the amino acid sequence of human MANCO 511
CC variant #1 transmembrane protein. The nucleic acid and polypeptide
CC sequences are useful for the diagnosis, prognosis and treatment of
CC immunological disorders (e.g. arthritis, graft rejection and acquired
CC immunodeficiency syndrome), inflammatory disorders (e.g. psoriasis and
CC asthma), renal disorders, embryonic disorders, brain-related disorders
CC (e.g. cerebral oedema), cerebrovascular diseases (e.g. ischaemia),
CC tumours, prostate-related disorders, pituitary-related disorders (e.g.
CC Cushing's disease) and neurodegenerative diseases (e.g. Parkinson's
CC disease).
XX
SQ Sequence 299 AA:

Query Match 13.8%; Score 91.5; DB 22; Length 299;
Best Local Similarity 26.2%; Pred. No. 0.012;
Matches 33; Conservative 18; Mismatches 48; Indels 27; Gaps 5;

QY 1 MSKLVVLLLMGVTWGPRVEAIFETQXSLMAE-----SEHXIKTLGQCDAD----48
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 19 vspalmvlllc]gslgprthvqgnlskaltwaegpsvlsrgnsvtlrcqglleaageyl 78

QY 49 -VPRPGDSRLRAVQEMGADPERVHLDSPAIKHQFLTLGTQO---GRYRCSGSLSTGMXOL 104
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 vkegsp-----epwdtqnpl---epknkarfslpsmlehnagryrcyyyspaqwapsep 127

QY 105 SKLLEL 110
: : : : :
Db 128 sdplel 133
```

```
RESULT 9
AAU01392
ID AAU01392 standard; Protein; 299 AA.
XX
AC AAU01392;
XX
DT 18-JUL-2001 (first entry)
XX
DE Human MANCO 511, variant #2 amino acid sequence.
XX
KW Human; MANCO 511; transmembrane protein; diagnostic; asthma;
KW immunological disorder; arthritis; graft rejection; renal disorder;
KW acquired immunodeficiency syndrome; inflammatory disorders; psoriasis;
KW AIDS; embryonic disorder; brain; cerebral oedema; ischaemia; tumour;
KW prostate; cerebrovascular disease; pituitary; Cushing's disease;
KW neurodegenerative disease; Parkinson's disease.
XX
OS Homo sapiens.
XX
PN WO200121631-A2.
XX
PD 29-MAR-2001.
XX
PF 20-SEP-2000; 2000WO-US25982.
XX
PR 20-SEP-1999; 99US-0399723.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Kirst SJ, Sharp JD, Fraser CC, Barnes T, Kingsbury G;
XX
DR MPI: 2001-211461/21.
XX
DR N-PSDB: AAS02103.
XX
PT New nucleic acid encoding INTERCEPT 307, MANCO 511, TANGO 351, TANGO
PT 361, TANGO 499 or TANGO 509 secreted or transmembrane protein, useful
PT for the diagnosis and treatment of arthritis, psoriasis and Parkinson's
PT disease -
XX
XX
PS Disclosure: Page 297-298; 362pp; English.
XX
XX
CC The sequence represents the amino acid sequence of human MANCO 511
CC variant #2 transmembrane protein. The nucleic acid and polypeptide
CC sequences are useful for the diagnosis, prognosis and treatment of
CC immunological disorders (e.g. arthritis, graft rejection and acquired
CC immunodeficiency syndrome), inflammatory disorders (e.g. psoriasis and
CC asthma), renal disorders, embryonic disorders, brain-related disorders
CC (e.g. cerebral oedema), cerebrovascular diseases (e.g. ischaemia),
CC tumours, prostate-related disorders, pituitary-related disorders (e.g.
CC Cushing's disease) and neurodegenerative diseases (e.g. Parkinson's
CC disease).
XX
SQ Sequence 299 AA:

Query Match 13.8%; Score 91.5; DB 22; Length 299;
Best Local Similarity 26.2%; Pred. No. 0.012;
Matches 33; Conservative 18; Mismatches 48; Indels 27; Gaps 5;

QY 1 MSKLVVLLLMGVTWGPRVEAIFETQXSLMAE-----SEHXIKTLGQCDAD----48
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 19 vspalmvlllc]gslgprthvqgnlskaltwaegpsvlsrgnsvtlrcqglleaageyl 78

QY 49 -VPRPGDSRLRAVQEMGADPERVHLDSPAIKHQFLTLGTQO---GRYRCSGSLSTGMXOL 104
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 vkegsp-----epwdtqnpl---epknkarfslpsmlehnagryrcyyyspaqwapsep 127

QY 105 SKLLEL 110
: : : : :
Db 128 sdplel 133
```



```

RESULT 10
AAW69233
ID AAW69233 standard; Protein; 623 AA.
XX
XX AAW69233;
XX
XX 20-OCT-1998 (first entry)
XX
XX
XX FcR-III protein sequence.
XX
XX Fc receptor-like protein; phagocytosis inducer; rheumatoid arthritis;
XX immune complex related disease; systemic lupus erythematosus; allergy;
XX haemolytic anaemia; thrombocytopenia; anaphylaxis; cancer; lymphoma;
XX leukaemia; infection; immunomodulator; viral entry inhibitor; therapy;
XX FcR-III.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 1..16
XX FT /note="signal peptide"
XX FT 17..623
XX FT /note="mature FcR-III"
XX
XX WO9831806-A2.
XX
XX 23-JUL-1998.
XX
XX 20-JAN-1998; 98WO-US01184.
XX
XX 18-JUN-1997; 97US-0049872.
XX PR 21-JAN-1997; 97US-0034205.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Gentz RL, Murphy M, NI J, Olsen HS, Ruben SM;
XX
XX MPI: 1998-414105/35.
XX DR N-PSDB; AAV44826.
XX
XX Nucleic acid encoding Fc receptor-like polypeptides or their
XX fragments - and related vectors, transformed cells and antibodies,
XX useful for treating and diagnosing diseases of the haematopoietic
XX and immune systems
XX
XX Claim 23; Fig 3A; 141pp; English.
XX
XX This sequence is the Fc receptor-like III protein (FcR-III) of the
XX invention. Cells containing the DNA are used to express the recombinant
XX protein, and to screen for specific (ant)agonists. The proteins are used
XX to induce phagocytosis, and their (ant)agonists are used to treat immune
XX complex related diseases (e.g. rheumatoid arthritis, systemic lupus
XX erythematosus, haemolytic anaemia, thrombocytopenia, anaphylaxis,
XX allergy, colorectal or breast cancer, lymphoma, leukaemia, infection by
XX intracellular pathogens etc). The antagonists are also useful as
XX immunomodulators and inhibitors of viral (e.g. human immune deficiency or
XX dengue viruses) entry into cells. The proteins may also be used to screen
XX for specific binding agents, i.e. (ant)agonists, for raising antibodies
XX (Ab), and for identification of particular cells or tissues. The Ab can
XX be used therapeutically as antagonists; as assay reagents for diagnostic
XX determination of the levels of expression of the proteins and for
XX affinity purification of the proteins. The DNA and its fragments are
XX useful as hybridisation probes or primers for isolating related genes, in
XX situ hybridisation (chromosome mapping) and diagnostically to measure
XX mRNA expression.
XX
XX Sequence 623 AA;

```

Query Match 13.7%; Score 90.5; DB 19; Length 623;
 Best Local Similarity 28.3%; Pred. No. 0.04;
 Matches 39; Conservative 10; Mismatches 38; Indels 51; Gaps 6;

```

OY 1 MSMLVFLILMGVTWGPVTEAIFETQXSLWAESHEHLKTLGGCCADVPDPPGDSRLPA 60
   | : | | | : | | | | |
Db 1 mtpalatlclclglsigrtrvgagpfpkpllwae-----pg-----s 37
OY 61 VOEWG-----AOE-----PVHLD-----SPAIRKHOFLTGDTQ---GRIR 92
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 38 viswgsprvliwcgsleaqeyqldekgspeidrnlnplekxkarfslpsmtqhagryr 97
OY 93 CRSGLSTGWXQXLSKLEL 110
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98 chyyssagwsepsdplrl 115

RESULT 11
AAW82552
ID AAW82552 standard; Protein; 631 AA.
XX
XX AAW82552;
XX
XX 09-FEB-1999 (first entry)
XX
XX Human LIR-pbm17 protein.
XX
XX LIR-pbm17; immunoregulator; leukocyte immunoglobulin-like receptor;
XX therapeutic; treatment; disorder; ITIM; MHC class I receptor; inhibitor;
XX cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
XX negative signalling; autoimmune disease; suppressor; LIR.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 1..16
XX FT /label="signal"
XX FT /label="extracellular_domain"
XX FT 17..631
XX FT /label="LIR-pbm17"
XX FT 444..464
XX FT /label="transmembrane_domain"
XX FT 465..631
XX FT /label="cytoplasmic_domain"
XX FT 514..517
XX FT /note="ITIM YxxL/V motif"
XX FT 543..546
XX FT /note="ITIM YxxL/V motif"
XX FT 595..598
XX FT /note="ITIM YxxL/V motif"
XX FT 625..628
XX FT /note="ITIM YxxL/V motif"
XX
XX WO9848017-A1.
XX
XX 29-OCT-1998.
XX
XX 23-APR-1998; 98WO-US08244.
XX
XX 24-APR-1997; 97US-0842248.
XX
XX (IMWV ) IMMUNEX CORP.
XX
XX Cosman DJ;
XX
XX MPI: 1998-609990/51.
XX DR N-PSDB; AAV69338.
XX
XX Leukocyte immunoglobulin-like receptor, LIR, polypeptides - useful,
XX e.g. for treating autoimmune diseases or disease states associated
XX with suppressed immune function
XX
XX Claim 4; Page 89-91; 112pp; English.
XX
XX This sequence represents a novel leukocyte immunoglobulin-like receptor
XX (LIR) polypeptide LIR-pbm17. This sequence can be administered

```

CC therapeutically to treat disorders associated with insufficient/defective
CC amounts of LIR polypeptide. LIR-P3G2 and certain other LIR family members
CC contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
CC (ITIMs), whilst other LIR family members lack ITIMs. By analogy with the
CC structure and function of known MHC Class I receptor molecules, LIRs
CC having ITIMs are inhibitory receptors mediating negative signalling,
CC whilst those lacking ITIMs are activatory receptors. Failure of a
CC receptor that mediates negative signalling could result in autoimmune
CC diseases, whilst failure of a receptor mediating activatory signalling
CC could result in suppressed immune function. They are also useful to
CC produce probes for detecting LIR nucleic acids or isolating LIR DNA from
CC other species.

XX Sequence 631 AA:

SQ

Query Match 13.7%: Score 90.5; DB 19; Length 631;
Best Local Similarity 28.3%: Pred. No. 0.041;
Matches 39; Conservative 10; Mismatches 38; Indels 51; Gaps 6;

OY 1 MSMLVYFLLMGVTGCPVTEAIFETQXSLMASESHXIKTLGQCDADVPGPGDSRLPA 60
DB 1 mcpalcalliclgisigprtrvgagpfpkplwae-----pg-----s 37
OY 61 VOEWG-----AOE-----PVHLD-----SPAIRKQFLTGDTQ---GRYR 92
DB 38 vlswgspvrlwcqgsleageyqldkgspepdlrnpkpkarlsfsmqghagryr 97
OY 93 CRSGLSTGKXQLSKLEL 110
DB 98 chysasagwsepsdplel 115

RESULT 12

AAB04177 standard; Protein: 631 AA.

XX AAB04177;

DT 11-APR-2001 (first entry)

DE Leukocyte immunoglobulin like receptor pblm17.

KW Leukocyte immunoglobulin like receptor; LIR; gene therapy;
KW autoimmunity; autoimmune disorders; immune system; human.

OS Homo sapiens.

PN WO200068383-A2.

PD 16-NOV-2000.

PF 12-MAY-2000: 2000MO-US13228.

PR 12-MAY-1999: 99US-0310463.

PA (IMMUNEX) IMMUNEX CORP.

PI Cosman DJ, Anderson DM, Borges L;

DR WPI: 2000-687445/67.

DR N-PSDB: AAA54601.

PT Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
PT tracing autoimmune diseases and disease states with suppressed immune
PT function

PS Example 9: Page 95-97; 117pp: English.

CC Host cells transformed with a vector which are capable of
CC expressing a leukocyte immunoglobulin like receptor (LIR), can be
CC used to produce the LIR polypeptide. LIR coding sequences may be
CC used in the gene therapy of disorders mediated directly or

CC indirectly by defective or insufficient amounts of any of the LIR
CC polypeptides. The LIR polypeptides can be used to treat autoimmune
CC diseases and disease states with suppressed immune function.

XX Sequence 631 AA:

SQ

Query Match 13.7%: Score 90.5; DB 21; Length 631;
Best Local Similarity 28.3%: Pred. No. 0.041;
Matches 39; Conservative 10; Mismatches 38; Indels 51; Gaps 6;

OY 1 MSMLVYFLLMGVTGCPVTEAIFETQXSLMASESHXIKTLGQCDADVPGPGDSRLPA 60
DB 1 mcpalcalliclgisigprtrvgagpfpkplwae-----pg-----s 37
OY 61 VOEWG-----AOE-----PVHLD-----SPAIRKQFLTGDTQ---GRYR 92
DB 38 vlswgspvrlwcqgsleageyqldkgspepdlrnpkpkarlsfsmqghagryr 97
OY 93 CRSGLSTGKXQLSKLEL 110
DB 98 chysasagwsepsdplel 115

RESULT 13

AAB61263 standard; Protein: 631 AA.

XX AAB61263;

DT 04-APR-2001 (first entry)

DE Human monocyte inhibitory receptor precursor.

KW Human; monocyte inhibitory receptor precursor; TANGO 268; cardiant;
KW cerebroprotective; cytoskeletal; anticoagulant; thrombolytic;
KW antiarteriosclerotic; haemostatic; glycoprotein VI; GPVI;
KW platelet membrane glycoprotein receptor; bleeding disorder;
KW blood vessel injury; thrombotic disorder; haemorrhagic disorder; stroke;
KW ischemia; cardiovascular disease; immunological disease; liver disorder;
KW cancer.

OS Homo sapiens.

PN WO200100810-A1.

PD 04-JAN-2001.

PF 30-JUN-2000: 2000MO-US18152.

PR 30-JUN-1999: 99US-0345468.

PR 06-DEC-1999: 99US-0454824.

PR 14-FEB-2000: 2000US-0503387.

PA (MILL-) MILLENNIUM PHARM INC.

PI Busfield SJ, Villalob J, Jandrot-Perrus M, Vainchenker W, Gill DS;

PI Qian MD, Kingsbury G;

DR WPI: 2001-080877/09.

DR N-PSDB: AAF29472.

PT New genes encoding human platelet-expressed collagen receptor,
PT glycoprotein VI, and its modulators, useful for preventing, treating
PT and diagnosing hemorrhagic disorders, thrombotic diseases and
PT immunological disorders -

PS Disclosure: Fig 4A-4B: 227pp: English.

CC The present sequence is given in a specification relating to an isolated
CC nucleic acid molecule encoding a platelet membrane glycoprotein receptor
CC glycoprotein VI (GPVI), also called TANGO 268. The GPVI polynucleotides
CC and polypeptides and their modulators, e.g. antisense nucleic acids,

CC ribozymes and antibodies, are useful for preventing, treating and
 CC diagnosing disorders associated with aberrant expression or activity of
 CC GPII. These disorders include bleeding disorders
 CC (e.g. thrombocytopenia), blood vessel injury, thrombotic disorders
 CC (e.g. thrombotic occlusion of the coronary arteries), haemorrhagic
 CC disorders, coronary artery and cerebral artery diseases (e.g. stroke and
 CC ischaemia), cardiovascular diseases (e.g. atherosclerosis and myocardial
 CC infarction), immunological diseases (e.g. platelet disorder) and
 CC embryonic liver disorders. Preferably they are used to prevent acute
 CC cardiac ischaemia following angioplasty and metastatic cancers,
 CC especially of the colon and liver.

XX Sequence 631 AA;

Query Match 13.2%; Score 87.5; DB 22; Length 631;

Best Local Similarity 27.5%; Pred. No. 0.093;

Matches 38; Conservative 11; Mismatches 38; Indels 51; Gaps 6;

OY 1 MSMLVFLIMGVTPTEAIFETQSLMAESENHXLKTLGGQDADVPQPGDSRLPA 60

DB 1 mpaitallclglstlgrtrvqagpfkplwae-----pg-----s 37

OY 61 VOEWG-----AOE-----PVHND-----SPAIKHOFLLTGDYQ---GRYR 92

DB 38 vlswgspvltwqsglsaeqeyrlckegspspdrlnplekpkartslpsmtehagryr 97

OY 93 CRSGSLTGKXQLSKLLEL 110

DB 98 chysagwsepsdpl 115

RESULT 14

AAW69234 ID AAW69234 standard; Protein; 472 AA.

XX AC AAW69234;

XX DT 20-OCT-1998 (first entry)

XX DE FcR-IV protein sequence.

XX Fc receptor-like protein; phagocytosis inducer; rheumatoid arthritis;
 KW immune complex related disease; systemic lupus erythematosus; allergy;
 KW haemolytic anaemia; thrombocytopenia; anaphylaxis; cancer; lymphoma;
 KW leukaemia; infection; immunomodulator; viral entry inhibitor; therapy;
 KW FcR-IV.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..16 /note= "signal peptide"

FT Protein 17..472 /note= "mature FcR-IV"

XX PN WO9831806-A2.

XX PD 23-JUL-1998.

XX PF 20-JAN-1998; 98WO-US01184.

XX PR 18-JUN-1997; 97US-0049872.

XX PR 21-JAN-1997; 97US-0034205.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Gentz RL, Murphy M, Ni J, Olsen HS, Ruben SM;

XX DR MPI. 1998-414105/35.

XX DR N-PSDB: AAW44827.

XX PT Nucleic acid encoding Fc receptor-like polypeptides or their

PT fragments - and related vectors, transformed cells and antibodies,
 PT useful for treating and diagnosing diseases of the hematopoietic
 PT and immune systems

PS Claim 23; Fig 4A: 141pp; English.

XX This sequence is the Fc receptor-like IV protein (FcR-IV) of the
 CC invention. Cells containing the DNA are used to express the recombinant
 CC protein, and to screen for specific (ant)agonists. The proteins are used
 CC to induce phagocytosis, and their (ant)agonists are used to treat immune
 CC complex related diseases (e.g. rheumatoid arthritis, systemic lupus
 CC erythematosus, haemolytic anaemia, thrombocytopenia, anaphylaxis,
 CC allergy, colorectal or breast cancer, lymphoma, leukaemia, infection by
 CC intracellular pathogens etc). The antagonists are also useful as
 CC immunomodulators and inhibitors of viral (e.g. human immune deficiency or
 CC dengue viruses) entry into cells. The proteins may also be used to screen
 CC for specific binding agents, i.e. (ant)agonists, for raising antibodies
 CC (Ab), and for identification of particular cells or tissues. The Ab can
 CC be used therapeutically as antagonists; as assay reagents for diagnostic
 CC determination of the levels of expression of the proteins and for
 CC affinity purification of the proteins. The DNA and its fragments are
 CC useful as hybridisation probes or primers for isolating related genes, in
 CC situ hybridisation (chromosome mapping) and diagnostically to measure
 CC mRNA expression.

XX Sequence 472 AA;

Query Match 12.9%; Score 85.5; DB 19; Length 472;

Best Local Similarity 27.9%; Pred. No. 0.11;

Matches 34; Conservative 11; Mismatches 44; Indels 33; Gaps 5;

OY 8 LILMGVTWGPVTEAIFETQSLMAE-----SEHXLKTLGGQDA-----DVPG 51

DB 8 lllclglstlgrtrvqagpfkplwaeqsgvlsagvstlwgqlaeayrlckeespa 67

OY 52 PGDSRLPAVQEWGADEPVHNDSPAIRKHOF---LLTGDYGRYCRSGSLTGKXQLSKL 108

DB 68 p-----wdrgnpl-----epkpkartslpsmtehagryrcyrypsvqsgsdpl 113

OY 109 EL 110

DB 114 el 115

RESULT 15

AAW82551 ID AAW82551 standard; Protein; 448 AA.

XX AC AAW82551;

XX DT 09-FEB-1999 (first entry)

XX DE Human LIR-pbm2 protein.

XX LIR-pbm2; immunoregulator; leukocyte immunoglobulin-like receptor;
 KW therapeutic treatment; disorder; ITIM; MHC class I receptor; inhibitor;
 KW cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
 KW negative signalling; autoimmune disease; suppressor; LIR.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..16 /label= signal

FT Domain 1..259 /label= extracellular_domain

FT Protein 17..448 /label= LIR-pmbh

FT Domain 260..280 /label= transmembrane_domain

FT Domain 281..448 /label= cytoplasmic_domain

```
FT Region 412..415
FT /note= "ITIM motif"
FT Region 442..445
FT /note= "ITIM motif"
PN WO9848017-A1.
XX
XX
XX 29-OCT-1998.
XX
XX 23-APR-1998; 98WO-US08244.
XX
XX 24-APR-1997; 97US-0842248.
XX
XX (IMM) ) IMMUNEX CORP.
XX
XX Cosman DJ.
XX
XX WPI: 1998-609990/51.
XX
XX N-PSDB: AAV69337.
XX
XX Leukocyte immunoglobulin-like receptor, LIR, polypeptides - useful,
XX e.g. for treating autoimmune diseases or disease states associated
XX with suppressed immune function
XX
XX Claim 4: Page 85-86; 112pp; English.
XX
XX This sequence represents a novel leukocyte immunoglobulin-like receptor
XX (LIR) polypeptide LIR-pbm2. This sequence can be administered
XX therapeutically to treat disorders associated with insufficient/defective
XX amounts of LIR polypeptide. LIR-p3g2 and certain other LIR family members
XX contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
XX (ITIMs), whilst other LIR family members lack ITIMs. By analogy with the
XX structure and function of known MHC Class I receptor molecules, LIRs
XX having ITIMs are inhibitory receptors mediating negative signalling,
XX whilst those lacking ITIMs are activatory receptors. Failure of a
XX receptor that mediates negative signalling could result in autoimmune
XX diseases, whilst failure of a receptor mediating activatory signalling
XX could result in suppressed immune function. They are also useful to
XX produce probes for detecting LIR nucleic acids or isolating LIR DNA from
XX other species.
XX
XX Sequence 448 AA:
SQ
Query Match 12.8%; Score 84.5; DB 19; Length 448;
Best Local Similarity 27.9%; Pred. No. 0.13;
Matches 34; Conservative 11; Mismatches 44; Indels 33; Gaps 5;
OY 8 LLLMGVTWGPVTEALFYETOXSLMAE-----SEHXKLTLCGCCDA-----DVPG 51
DB 8 11c1g1s1gprcthmgaqg1pkp1lwepgsv1sgnsv1wcg1learey1dkeespa 67
OY 52 PPGDSRLPAVQEMGAEPVHNLSPAIKHOF---LITGDTQGRYRCHSG1STGKXOLSKLL 108
DB 68 p-----wdtqnpl-----epknkartsipmledyagrycyrgspvsgsqpsdpl 113
OY 109 EL 110
DB 114 el 115
RESULT 16
AAW53463
ID AAW53463 standard; Protein; 448 AA.
XX
XX AAW53463;
XX
XX 17-JUL-1998 (first entry)
XX
XX Human gp49 HM18 polypeptide.
XX
XX Human; gp49; HM18; HM43; Immunoglobulin; Immune response; mast cell;
XX bone marrow; cell-surface member; FCERI.
XX
XX
```

```
XX
XX Homo sapiens.
XX
XX OS WO9809638-A1.
XX
XX PN 12-MAR-1998.
XX
XX PD 05-SEP-1997; 97WO-US15586.
XX
XX PF 06-SEP-1996; 96US-0025846.
XX
XX PR (BCHM ) BRIGHAM & WOMENS HOSPITAL.
XX
XX PA Arm JP, Austen KF, Castellis MC, Katz HR;
XX
XX PI WPI: 1998-193318/17.
XX
XX DR N-PSDB: AAV23273.
XX
XX Cell-surface member of immunoglobulin super-family, human gp49 -
XX useful to treat undesired immune responses, especially mast
XX cell-related diseases
XX
XX Claim 5: Fig 1A-B; 62pp; English.
XX
XX The present sequence represents human gp49 polypeptide HM18. The present
XX invention also describes: (1) a fusion polypeptide comprising a human
XX gp49 component and FCERI or a detectable marker; (2) a recombinant
XX nucleic acid encoding human gp49 or a human gp49-FCERI fusion as above;
XX (3) a cell or vector comprising the recombinant nucleic acid as in (2),
XX and (4) an antibody which selectively binds to gp49. Mammalian gp49 or
XX its related DNA can be used to treat an undesired immune response,
XX especially a mast cell-related disease.
XX
XX Sequence 448 AA:
SQ
Query Match 12.8%; Score 84.5; DB 19; Length 448;
Best Local Similarity 27.9%; Pred. No. 0.13;
Matches 34; Conservative 11; Mismatches 44; Indels 33; Gaps 5;
OY 8 LLLMGVTWGPVTEALFYETOXSLMAE-----SEHXKLTLCGCCDA-----DVPG 51
DB 8 11c1g1s1gprcthmgaqg1pkp1lwepgsv1sgnsv1wcg1learey1dkeespa 67
OY 52 PPGDSRLPAVQEMGAEPVHNLSPAIKHOF---LITGDTQGRYRCHSG1STGKXOLSKLL 108
DB 68 p-----wdtqnpl-----epknkartsipmledyagrycyrgspvsgsqpsdpl 113
OY 109 EL 110
DB 114 el 115
RESULT 17
AAB04176
ID AAB04176 standard; Protein; 448 AA.
XX
XX AAB04176;
XX
XX 11-APR-2001 (first entry)
XX
XX Leukocyte immunoglobulin like receptor pbm2.
XX
XX Leukocyte immunoglobulin like receptor; LIR; gene therapy;
XX autoimmunity; autoimmune disorders; immune system; human.
XX
XX Homo sapiens.
XX
XX OS WO200068383-A2.
XX
XX PN 16-NOV-2000.
XX
XX PD 12-MAY-2000; 2000WO-US13228.
XX
XX
```

```

XX 12-MAY-1999; 99US-0310463.
PR XX
XX (IMMV ) IMMUNEX CORP.
PA XX
XX Cosman DJ, Anderson DM, Borges L;
XX WPI: 2000-687645/67.
XX N-PSDB; AAA54600.
XX
XX Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
XX treating autoimmune diseases and disease states with suppressed immune
XX function
XX
XX Example 9; Page 91-92; 117pp; English.
XX
XX Host cells transformed with a vector which are capable of
XX expressing a leukocyte immunoglobulin like receptor (LIR), can be
XX used to produce the LIR polypeptide. LIR coding sequences may be
XX used in the gene therapy of disorders mediated directly or
XX indirectly by defective or insufficient amounts of any of the LIR
XX polypeptides. The LIR polypeptides can be used to treat autoimmune
XX diseases and disease states with suppressed immune function.
XX
XX Sequence 448 AA:
SQ
Query Match 12.8%; Score 84.5; DB 21; Length 448;
Best Local Similarity 27.9%; Pred. No. 0.13;
Matches 34; Conservative 11; Mismatches 44; Indels 33; Gaps 5;
QY 8 LLLMGVTMGPRVPEAIFETQXSLMAE-----SEHXIKTLTGQCD-----DVPG 51
DB 8 LLLGISTISPRTHMGAGPLPKPLWAEPGSVLSWNSVTLWCQGLTAREYRLDKESPA 67
QY 52 PGDSRLPAVQEWGAQEPVHLDSPAIKHOF---LLTGDTQGRYRCRSGSLSTGWXQLSKL 108
DB 68 P-----wdqgnpl---epkkrkfsspsmtefyagrycyrsyrgwsgpsdpl 113
QY 109 EL 110
DB 114 el 115

```

RESULT 18
AAW82553
ID AAW82553 standard; Protein: 590 AA.
XX
XX AAW82553;
AC XX
XX 09-FEB-1999 (first entry)
DT XX
XX Human LIR-pbmnew protein.
DE XX
XX LIR-pbmnew; immunoregulator; leukocyte immunoglobulin-like receptor;
KW therapeutic; treatment; disorder; ITIM; MHC class I receptor; inhibitor;
KW cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
KW negative signaling; autoimmune disease; suppressor; LIR.
XX
XX Homo sapiens.
OS XX
XX
XX Key Location/Qualifiers
FH 1..16 /label= signal
FT 1..456 /label= extracellular_domain
FT 17..590 /label= LIR-pbmnew
FT 457..579 /label= transmembrane_domain
FT 580..590 /label= cytoplasmic_domain
FT 554..557 /label= Region

```

FT FT /note="ITIM motif"
FT Region 584..587
FT FT /note="ITIM motif"
XX
XX WO9848017-A1.
XX
XX 29-OCT-1998.
XX
XX 23-APR-1998; 98WO-US08244.
XX
XX 24-APR-1997; 97US-0842248.
XX
XX (IMMV ) IMMUNEX CORP.
XX
XX Cosman DJ;
XX
XX WPI: 1998-609990/51.
XX N-PSDB; AAV69339.
XX
XX Leukocyte immunoglobulin-like receptor, LIR, polypeptides - useful,
XX e.g. for treating autoimmune diseases or disease states associated
XX with suppressed immune function
XX
XX Claim 4; Page 94-96; 112pp; English.
XX
XX This sequence represents a novel leukocyte immunoglobulin-like receptor
XX (LIR) polypeptide LIR-pbmnew. This sequence can be administered
XX therapeutically to treat disorders associated with insufficient/defective
XX amounts of LIR polypeptide. LIR-p302 and certain other LIR family members
XX contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
XX (ITIMs), whilst other LIR family members lack ITIMs. By analogy with the
XX structure and function of known MHC Class I receptor molecules, LIRs
XX having ITIMs are inhibitory receptors mediating negative signaling,
XX whilst those lacking ITIMs are activatory receptors. Failure of a
XX receptor that mediates negative signaling could result in autoimmune
XX diseases, whilst failure of a receptor mediating activatory signaling
XX could result in suppressed immune function. They are also useful to
XX produce probes for detecting LIR nucleic acids or isolating LIR DNA from
XX other species.
XX
XX Sequence 590 AA:
SQ
Query Match 12.3%; Score 81.5; DB 19; Length 590;
Best Local Similarity 27.9%; Pred. No. 0.44;
Matches 38; Conservative 16; Mismatches 55; Indels 27; Gaps 6;
QY 1 MSMLVFLILMGVTMGPRVPEAIFETQXSLMAESEHXLK-----TL-----GQCD----- 47
DB 1 MLLTSLVLLICGLSVPTCVGAGPLPKPLWAEPASVLSWNSVTLWCQGLTAREYRL 60
QY 48 DVPGPDSRLPAVQEWGAQEPVHLDSPAIKHOFLLTGDTQGRYRCRSGSLSTGWXQLSKL 107
DB 61 dkeglpwarkrpnrlpegakfhpslty-----daagryrcyretpragwsepsdp 112
QY 108 LEL--TG----PKVLA 117
DB 113 lslvatglfyaeptlla 128

```

RESULT 19
AAB04178
ID AAB04178 standard; Protein: 590 AA.
XX
XX AAB04178;
AC XX
XX 11-APR-2001 (first entry)
DT XX
XX Leukocyte immunoglobulin like receptor pbmnew.
DE XX
XX Leukocyte immunoglobulin like receptor; LIR; gene therapy;
KW autoimmunity; autoimmune disorders; immune system; human.
XX

OS Homo sapiens.
XX MO200068383-A2.
XX 16-NOV-2000.
XX 12-MAY-2000; 2000MO-US13228.
XX 12-MAY-1999; 99US-0310463.
XX (IMMUNEX CORP.
XX Cosman DJ, Anderson DM, Borges L;
XX WPI: 2000-687645/67.
XX N-PSDB: AAB04178.
XX
XX Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
XX treating autoimmune diseases and disease states with suppressed immune
XX function
XX
XX Disclosure: Page 100-101; 117pp: English.
XX
XX Host cells transformed with a vector which are capable of
XX expressing a leukocyte immunoglobulin like receptor (LIR), can be
XX used to produce the LIR polypeptide. LIR coding sequences may be
XX used in the gene therapy of disorders mediated directly or
XX indirectly by defective or insufficient amounts of any of the LIR
XX polypeptides. The LIR polypeptides can be used to treat autoimmune
XX diseases and disease states with suppressed immune function.
XX
XX Sequence 590 AA:

Query Match 12.3%; Score 81.5; DB 21; Length 590;
Best Local Similarity 27.9%; Pred. No. 0.44;
Matches 38; Conservative 16; Mismatches 55; Indels 27; Gaps 6;
QY 1 MSMLVYFLLMGVTGWPVBEAIFETQXSLMAESEHXLK-----TL---GCOCDA-----47
DB 1 mltlsvllclglsvgrptcvgqglpkrlwepasvlarqkpvllwcqgleteyr1 60
QY 48 DVPGPCDSRLPAVQDGAQEPVHLDSPAIKHQFLTCTPGTCGRYRCRSGSLTCGXOLSK1 107
DB 61 dkeglpwmkrnpqlepgkakfhlpstvy-----dsagryrcryclcpagwspsdp 112
QY 108 LEL--TG----PKVLA 117
DB 113 lclvatgfyaeplla 128

RESULT 20
AAM82545
ID AAM82545 standard; Protein; 652 AA.
XX
XX AAM82545;
XX
XX 09-FEB-1999 (first entry)
XX
XX Human LIR-18A3 protein.
XX
XX LIR-18A3; immunoregulator; leukocyte immunoglobulin-like receptor; LIR;
XX therapeutic; treatment; disorder; ITIM; MHC class I receptor; inhibitor;
XX cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
XX negative signalling; autoimmune disease; suppressor.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..16
XX FT /label signal
XX FT 1..459
XX FT Domain /label extracellular_domain

FT Protein 17..652
FT /label LIR-18A3
FT Domain 460..484
FT /label transmembrane_domain
FT Domain 485..652
FT /label cytoplasmic_domain
FT Region 534..537
FT /note "ITIM motif"
FT Region 564..567
FT /note "ITIM motif"
FT Region 616..619
FT /note "ITIM motif"
FT Region 646..649
FT /note "ITIM motif"
XX
XX W09848017-A1.
XX
XX 29-OCT-1998.
XX
XX 23-APR-1998; 98WO-US08244.
XX
XX 24-APR-1997; 97US-0842248.
XX
XX (IMMUNEX CORP.
XX
XX Cosman DJ;
XX
XX WPI: 1998-609990/51.
XX N-PSDB: AAV69329.
XX
XX Leukocyte immunoglobulin-like receptor, LIR, polypeptides - useful,
XX e.g. for treating autoimmune diseases or disease states associated
XX with suppressed immune function
XX
XX Claim 4; Page 59-61; 112pp: English.
XX
XX This sequence represents a novel leukocyte immunoglobulin-like receptor
XX (LIR) polypeptide LIR-18A3. This sequence can be administered
XX therapeutically to treat disorders associated with insufficient/defective
XX amounts of LIR polypeptide. LIR-P3G2 and certain other LIR family members
XX contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
XX (ITIMs), whilst other LIR family members lack ITIMs. By analogy with the
XX structure and function of known MHC Class I receptor molecules, LIRs
XX having ITIMs are inhibitory receptors mediating negative signalling,
XX whilst those lacking ITIMs are activatory receptors. Failure of a
XX receptor that mediates negative signalling could result in autoimmune
XX diseases, whilst failure of a receptor mediating activatory signalling
XX could result in suppressed immune function. They are also useful to
XX produce probes for detecting LIR nucleic acids or isolating LIR DNA from
XX other species.
XX
XX Sequence 652 AA:

Query Match 11.9%; Score 78.5; DB 19; Length 652;
Best Local Similarity 24.6%; Pred. No. 1.1;
Matches 41; Conservative 16; Mismatches 57; Indels 53; Gaps 7;
QY 1 MSMLVYFLLMGVTGWPVBEAIFETQXSLMAESEHXLKTLCCACADVPGPCDSRLPA 60
DB 1 mltlsvllclglsvgrptcvgqglpkrlwepasvlarqkpvllwcqgleteyr1 60
QY 61 VQE-----WGAEQPVHLDSPAIKHQF---LITGTCGRYRCRSGSLT-GMX 102
DB 52 gqetqeyrlryekktalwrlrtpqel---vkxqqrpsltwheagryrcygsdtagrs 108
QY 103 QLSKLELT-----GPKVLAC--SLALDGAS 126
DB 109 esedplelvtvgaylkrpltsaqpsrvvnsgnvlbgcdqvaldgfs 155

RESULT 21
AAB04170

```

ID AAB04170 standard; Protein: 652 AA.
XX
AC AAB04170;
XX
DT 11-APR-2001 (first entry)
XX
DE Leukocyte immunoglobulin like receptor (LIR) 18A3.
XX
KM Leukocyte immunoglobulin like receptor; LIR: gene therapy;
XX autoimmunity; autoimmune disorders; Immune system; human.
XX
OS Homo sapiens.
XX
PN MO20068383-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000MO-US13228.
XX
PR 12-MAY-1999; 99US-0310463.
XX
PA (IMMV) IMMUNEX CORP.
XX
PI Cosman DJ, Anderson DM, Borges L;
XX
DR MPI: 2000-687645/67.
XX
DR N-PSDB: AAA54592.
XX
PT Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
XX treating autoimmune diseases and disease states with suppressed immune
XX function
XX
PS Example 3; Page 66-68; 117pp; English.
XX
CC Host cells transformed with a vector which are capable of
XX expressing a leukocyte immunoglobulin like receptor (LIR), can be
XX used to produce the LIR polypeptide. LIR coding sequences may be
XX CC in the gene therapy of disorders mediated directly or
XX indirectly by defective or insufficient amounts of any of the LIR
XX CC polypeptides. The LIR polypeptides can be used to treat autoimmune
XX diseases and disease states with suppressed immune function.
XX
SQ Sequence 652 AA:

Query Match 11.9%; Score 78.5; DB 21; Length 652;
Best Local Similarity 24.6%; Pred. No. 1.1;
Matches 41; Conservative 16; Mismatches 57; Indels 53; Gaps 7;

QY 1 MSMLVYFLIMGTWGPVTEAIFETQXSLMAESEHXLKTLGOCADADVGPFGDSKRLPA 60
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 1 mepdlvllcigslgprthvqgghlprkplwaepgsvl-----lqspvltlrcqg 51
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 61 VDE-----WGAQEPVHCDSPAIRKHQF---LITGPTGGRYRCRGLST-GWX 102
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 52 ggetggrlryrkktalwlttrlpge---vkxgqfipslitwahaqrryrysgdsdagrs 108
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 103 QLSKLLLELT-----GPRVLAC--SLALDGAS 126
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 109 essdpelvtvgaylkrplsaqpsrvnsgnvlldcdsqvaldgsf 155
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 22
AA57323
ID AAY57323 standard; Protein: 244 AA.
XX
AC AAY57323;
XX
DT 13-JUN-2000 (first entry)
XX
DE Nostoc NspI restriction endonuclease (NspIR).
XX
KW NspI restriction endonuclease; NspIR; Nostoc; NspI methylase; NspIrm.

```

```

XX
OS Nostoc sp.
XX
PN US6027929-A.
XX
PD 22-FEB-2000.
XX
PF 18-AUG-1998; 98US-0135782.
XX
PR 18-AUG-1998; 98US-0135782.
XX
PA (NEWE) NEW ENGLAND BIOLABS INC.
XX
PI Xu S;
XX
DR MPI: 2000-316238/27.
XX
DR N-PSDB: AA290654.
XX
PT Isolated DNA encoding the NspI endonuclease and methylase, useful for
XX creating recombinant DNA molecules -
XX
PS Claim 1; Fig 3; 14pp; English.
XX
CC The invention relates to a NspI restriction endonuclease (NspIR),
XX CC obtained from Nostoc sp. The NspIR is recombinantly obtained using a
XX CC NspI methylase (NspIm) premordified E. coli K strain RRI (gammade3).
XX CC The endonuclease and methylase are useful for creating recombinant DNA
XX CC molecules. The present sequence represents a Nostoc NspIR.
XX
SQ Sequence 244 AA:

Query Match 11.4%; Score 75.5; DB 21; Length 244;
Best Local Similarity 39.3%; Pred. No. 0.72;
Matches 22; Conservative 5; Mismatches 18; Indels 11; Gaps 4;

QY 55 DSRIPAVQENGAQDPVILDS---PAIRKHQF-LITGPTGGRYR-----CRSGLSTG 100
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 132 dhtrf-mewgaasephltsmdneikrkfqlkkdtsghmlkrsccerciktg 186
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 23
AA24321
ID AAY24321 standard; Protein: 635 AA.
XX
AC AAY24321;
XX
DT 16-SEP-1999 (first entry)
XX
DE Mouse dephosphorylase inhibiting pol-like protein #4.
XX
KW Dephosphorylase inhibiting protein; p91; tyrosine phosphatase SHP-1;
XX SHP-2; inositol-5-phosphate SHP; phosphorylating tyrosine;
XX immunoreceptor; immunomodulatory agent.
XX
OS Mus sp.
XX
PN JP1169184-A.
XX
PD 29-JUN-1999.
XX
PF 12-DEC-1997; 97JP-0362285.
XX
PR 12-DEC-1997; 97JP-0362285.
XX
PA (UYOK-) UNIV OKAYAMA.
XX
DR MPI: 1999-422622/36.
XX
DR N-PSDB: AAX88978.
XX
PT New peptide - useful for inhibiting dephosphorylase
XX
PS Claim 2; Page 20-21; 30pp; Japanese.

```


KM cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
 KW negative signalling; autoimmune disease; suppressor.

OS Homo sapiens.

XX Key Location/Qualifiers
 FH 1..16
 FT Peptide /label_ signal

FT Domain 1..261

FT Protein /label_ extracellular_domain

FT Domain 17..289

FT Domain /label_ LIR-pmb36-2

FT Domain 262..280

FT Domain /label_ transmembrane_domain

FT Domain 281..289

FT Domain /label_ cytoplasmic_domain

PN WO9848017-A1.

PD 29-OCT-1998.

PF 23-APR-1998: 98WO-US08244.

PR 24-APR-1997: 97US-0842248.

PA (IMMUNEX CORP.

PI Cosman DJ;

DR WPI: 1998-609990/51.

DR N-PSDB: AAV69334.

XX Leukocyte immunoglobulin-like receptor, LIR, polypeptides - useful,
 PT e.g. for treating autoimmune diseases or disease states associated
 with suppressed immune function

PS Claim 4: Page 73-74; 112pp; English.

XX This sequence represents a novel leukocyte immunoglobulin-like receptor
 CC (LIR) polypeptide LIR-pmb36-2. This sequence can be administered
 CC therapeutically to treat disorders associated with insufficient/defective
 CC amounts of LIR polypeptide. LIR-p362 and certain other LIR family members
 CC contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
 CC (ITIMs), whilst other LIR family members lack ITIMs. By analogy with the
 CC structure and function of known MHC class I receptor molecules, LIRs
 CC having ITIMs are inhibitory receptors mediating negative signalling,
 CC whilst those lacking ITIMs are activatory receptors. Failure of a
 CC receptor that mediates negative signalling could result in autoimmune
 CC diseases, whilst failure of a receptor mediating activatory signalling
 CC could result in suppressed immune function. They are also useful to
 CC produce probes for detecting LIR nucleic acids or isolating LIR DNA from
 CC other species.

XX Sequence 289 AA:

Query Match 11.2%; Score 74; DB 19; Length 289;

Best Local Similarity 27.7%; Pred. No. 1.4; Mismatches 45; Indels 34; Gaps 7;

Matches 36; Conservative 15; Mismatches 45; Indels 34; Gaps 7;

QY 1 MSMLVFLLMGVTGMPVTEAIFETQXSLMAESHXLTLCQ-----CDA----- 47

DB 1 mPpILVILICIGISgprchvgagLpKpLlwaepsvl-tqgsprvLwcgjlleqeyr 59

QY 48 ---DVGPPGDSRLPAVQEWGAQEPVNLDSPAIKHQF---LITGDTQGRYRCRSGLST-G 100

DB 60 lyreKktapwltirp-----geIvk-----kgqfipsltwentgyrctygshtag 106

QY 101 WXQLSKLLEL 110

DB 107 wepsdpl 116

RESULT 29

AAB04173

XX AAB04173 standard; Protein: 289 AA.

XX AAB04173;

DT 11-APR-2001 (first entry)

XX Leukocyte immunoglobulin like receptor pmb36-2.

XX Leukocyte immunoglobulin like receptor; LIR; gene therapy;

XX autoimmunity; autoimmune disorders; immune system; human; ds.

OS Homo sapiens.

PN WO200068383-A2.

PD 16-NOV-2000.

PF 12-MAY-2000: 2000WO-US13228.

PR 12-MAY-1999: 99US-0310463.

PA (IMMUNEX CORP.

PI Cosman DJ, Anderson DM, Borges L;

DR WPI: 2000-687645/67.

DR N-PSDB: AAA54597.

XX Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
 PT treating autoimmune diseases and disease states with suppressed immune
 PT function

PS Example 9; Page 79-80; 117pp; English.

XX Host cells transformed with a vector which are capable of
 CC expressing a leukocyte immunoglobulin like receptor (LIR), can be
 CC used to produce the LIR polypeptide. LIR coding sequences may be
 CC used in the gene therapy of disorders mediated directly or
 CC indirectly by defective or insufficient amounts of any of the LIR
 CC polypeptides. The LIR polypeptides can be used to treat autoimmune
 CC diseases and disease states with suppressed immune function.

XX Sequence 289 AA:

Query Match 11.2%; Score 74; DB 21; Length 289;

Best Local Similarity 27.7%; Pred. No. 1.4; Mismatches 45; Indels 34; Gaps 7;

Matches 36; Conservative 15; Mismatches 45; Indels 34; Gaps 7;

QY 1 MSMLVFLLMGVTGMPVTEAIFETQXSLMAESHXLTLCQ-----CDA----- 47

DB 1 mPpILVILICIGISgprchvgagLpKpLlwaepsvl-tqgsprvLwcgjlleqeyr 59

QY 48 ---DVGPPGDSRLPAVQEWGAQEPVNLDSPAIKHQF---LITGDTQGRYRCRSGLST-G 100

DB 60 lyreKktapwltirp-----geIvk-----kgqfipsltwentgyrctygshtag 106

QY 101 WXQLSKLLEL 110

DB 107 wepsdpl 116

RESULT 30

AAW82544

ID AAW82544 standard; Protein: 650 AA.

XX AAW82544;

XX 09-FEB-1999 (first entry)

XX Human LIR-P3G2 protein.

```

XX LIR-P3G2; leukocyte immunoglobulin-like receptor; LIR;
KW therapeutic; treatment; disorder; ITIM; MHC class I receptor; inhibitor;
KW cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
KW negative signaling; autoimmune disease; suppressor.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..16
FT /label= signal
FT Domain 1..458
FT /label= extracellular_domain
FT Protein 17..650
FT /label= LIR-P3G2
FT Domain 459..483
FT /label= transmembrane_domain
FT Domain 17..118
FT /note= "immunoglobulin-like domain I"
FT Domain 119..220
FT /note= "immunoglobulin-like domain II"
FT Domain 221..318
FT /note= "immunoglobulin-like domain III"
FT Domain 319..419
FT /note= "immunoglobulin-like domain IV"
FT Domain 484..650
FT /label= cytoplasmic_domain
FT Region 533..536
FT /note= "ITIM motif"
FT Region 562..565
FT /note= "ITIM motif"
FT Region 614..617
FT /note= "ITIM motif"
FT Region 644..647
FT /note= "ITIM motif"
XX
XX WO9848017-A1.
XX
XX 29-OCT-1998.
XX
XX 23-APR-1998; 98WO-US08244.
XX
XX 24-APR-1997; 97US-0842248.
XX
XX (IMMV ) IMMUNEX CORP.
XX
XX Cosman DJ;
XX
XX WPI: 1998-609990/51.
XX
XX N-PSDB; AAV69328.
XX
XX Leukocyte immunoglobulin-like receptor, LIR, polypeptides - useful,
XX e.g. for treating autoimmune diseases or disease states associated
XX with suppressed immune function
XX
XX Claim 4: Page 53-55; 112pp; English.
XX
XX This sequence represents a novel leukocyte immunoglobulin-like receptor
XX (LIR) polypeptide LIR-P3G2. This sequence can be administered
XX therapeutically to treat disorders associated with insufficient/defective
XX amounts of LIR polypeptide. LIR-P3G2 and certain other LIR family members
XX contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
XX (ITIMs). Whilst other LIR family members lack ITIMs. By analogy with the
XX structure and function of known MHC Class I receptor molecules, LIRs
XX having ITIMs are inhibitory receptors mediating negative signaling,
XX whilst those lacking ITIMs are activatory receptors. Failure of a
XX receptor that mediates negative signaling could result in autoimmune
XX diseases. Whilst failure of a receptor mediating activatory signaling
XX could result in suppressed immune function. They are also useful to
XX produce probes for detecting LIR nucleic acids or isolating LIR DNA from
XX other species.
XX
XX Sequence 650 AA;

```

```

Query Match 11.2%; Score 74; DB 19; Length 650;
Best Local Similarity 26.6%; Pred. No. 3.9;
Matches 34; Conservative 14; Mismatches 50; Indels 30; Gaps 5;

OY 1 MSMLVFLIMGVTVGVTPEAAIFETQXSLMAESEHXLKTLGOCDAVDGPGDSRLPA 60
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1 mcpilvlliclglsigrthvqagshlpkplwaegsvl-----tqgsrvllrcgq 51

OY 61 VDE-----WGAGEPVLHSDPAIKHQF---LLTGDYGRYRCRSGLST-GMX 102
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 52 gqetqeyrlrrekktpawtrirpgei---vkkqgfpdpsltwehagryrcyysdlaags 108

OY 103 QLSKLEL 110
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 109 essdplel 116

RESULT 31
AAB04169
ID AAB04169 standard; Protein: 650 AA.
AC AAB04169;
XX
XX 11-APR-2001 (first entry)
XX
XX Leukocyte immunoglobulin like receptor (LIR) P3G2.
XX
XX Leukocyte immunoglobulin like receptor; LIR; gene therapy;
XX autoimmunity; autoimmune disorders; immune system; human.
XX
XX OS Homo sapiens.
XX
XX WO200068383-A2.
XX
XX 16-NOV-2000.
XX
XX 12-MAY-2000; 2000WO-US13228.
XX
XX 12-MAY-1999; 99US-0310463.
XX
XX (IMMV ) IMMUNEX CORP.
XX
XX Cosman DJ, Anderson DM, Borges L;
XX
XX WPI: 2000-687645/67.
XX
XX N-PSDB; AAA54591.
XX
XX Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
XX treating autoimmune diseases and disease states with suppressed immune
XX function
XX
XX Example 3: Page 60-62; 117pp; English.
XX
XX Host cells transformed with a vector which are capable of
XX expressing a leukocyte immunoglobulin like receptor (LIR), can be
XX used to produce the LIR polypeptide. LIR coding sequences may be
XX used in the gene therapy of disorders mediated directly or
XX indirectly by defective or insufficient amounts of any of the LIR
XX polypeptides. The LIR polypeptides can be used to treat autoimmune
XX diseases and disease states with suppressed immune function.
XX
XX Sequence 650 AA;

Query Match 11.2%; Score 74; DB 21; Length 650;
Best Local Similarity 26.6%; Pred. No. 3.9;
Matches 34; Conservative 14; Mismatches 50; Indels 30; Gaps 5;

OY 1 MSMLVFLIMGVTVGVTPEAAIFETQXSLMAESEHXLKTLGOCDAVDGPGDSRLPA 60
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1 mcpilvlliclglsigrthvqagshlpkplwaegsvl-----tqgsrvllrcgq 51

```


CC to integrin alpha-v-beta-5 and includes a part of the C-terminal domain
CC of MMP-2. The antagonists are used to inhibit angiogenesis in inflamed
CC tissue, in solid tumours or metastases, and in a wide range of ocular
CC disorders (e.g. diabetic or other forms of retinopathy, neovascular
CC glaucoma, or corneal transplants). They are particularly used to induce
CC regression or to inhibit growth of tumours. The alpha-v-beta-5
CC antagonists can also be used to treat restenosis caused by migration of
CC smooth muscle cells following angioplasty and to reduce blood supply to
CC selected tissues. The antagonists particularly inhibit neovascularisation
CC where this is induced by cytokines, e.g. transforming growth factor
CC alpha, epidermal growth factor or especially vascular endothelial growth
CC factor. note: this sequence does not appear in the specification; it was
CC created using information provided.

SQ Sequence 228 AA:

Query Match 11.0%; Score 73; DB 19; Length 228;

Best Local Similarity 23.6%; Pred. No. 1.3;

Matches 26; Conservative 13; Mismatches 37; Indels 34; Gaps 4;

OY 17 PYTEAIFETQXSLMAESEHMLKTLGQCDADVPG-----PPGDSRLPAVQEWGACQEP 69

DB 121 pqdekavff-----agneywyvtasndrygpkltstgldpvdvqrdaafnwgrrn-- 171

OY 70 VHLDSPAIRHOFLLTGDRGRRCRSGSLTGWQXLSKLELTGPKVLACS 119

DB 172 -----kktylfsgdrywkyh-----eekkmelatpkrtiads 203

RESULT 38

AAW41111

ID AAW41111 standard; Protein; 663 AA.

XX AAW41111;

XX 08-JUN-1998 (first entry)

XX Chicken matrix metalloproteinase-2.

XX Matrix metalloproteinase-2; MMP-2; chicken;

XX Angiogenesis; inhibitor; antagonist; integrin alpha-v beta-3;

XX vitronectin receptor; rheumatoid arthritis; tumour; metastasis;

XX diabetic retinopathy; macular degeneration; restenosis; therapy.

XX Gallus sp.

XX Key Location/Qualifiers

XX Peptide 1..26

XX /label= Sig-peptide

XX WO9745137-A1.

XX 04-DEC-1997.

XX 30-MAY-1997; 97WO-US09158.

XX 31-MAY-1996; 96US-0018733.

XX 31-MAY-1996; 96US-0015869.

XX (SCRI) SCRIPPS RES INST.

XX Brooks P, Cheresh DA;

XX WPI; 1998-032334/03.

XX N-PSDB; AAV03995.

XX Packaging material containing polypeptide antagonist of alphav,

XX betas integrin - used for inhibition of angiogenesis, and for

XX treating tumours, inflammation, eye diseases etc.

XX Disclosure: Page 163-167, 234pp: English.

CC This protein sequence comprises chicken matrix metalloproteinase-2
CC (chMMP-2). The invention relates to the discovery that angiogenesis
CC is mediated by the specific vitronectin receptor alpha-v beta-3,
CC and that inhibition of alpha-v beta-3 function inhibits
CC angiogenesis. Claimed antagonists of alpha-v beta-3 include
CC C-terminal fragments (see AAW41083-94) of human or chicken MMP-2. An
CC MMP-2 fragment can be obtained by recombinant DNA methods, such as
CC PCR amplification of the chMMP-2 coding region, cloning into e.g.
CC pGEX-3X, and expression in E. coli as a fusion protein with
CC glutathione-S-transferases. The antagonists can be used to inhibit
CC angiogenesis in inflamed tissue (for treatment of arthritis or
CC rheumatoid arthritis), in solid tumours or metastases (particularly
CC to induce regression or inhibit tumour growth), and in ocular
CC disorders such as diabetic retinopathy and macular degeneration, as
CC well as to treat restenosis (all claimed).

SQ Sequence 663 AA:

Query Match 11.0%; Score 73; DB 19; Length 663;

Best Local Similarity 23.6%; Pred. No. 5.3;

Matches 26; Conservative 13; Mismatches 37; Indels 34; Gaps 4;

OY 17 PYTEAIFETQXSLMAESEHMLKTLGQCDADVPG-----PPGDSRLPAVQEWGACQEP 69

DB 530 pqdekavff-----agneywyvtasndrygpkltstgldpvdvqrdaafnwgrrn-- 580

OY 70 VHLDSPAIRHOFLLTGDRGRRCRSGSLTGWQXLSKLELTGPKVLACS 119

DB 581 -----kktylfsgdrywkyh-----eekkmelatpkrtiads 612

RESULT 39

AAW41227

ID AAW41227 standard; protein; 663 AA.

XX AAW41227;

XX 09-JUN-1998 (first entry)

XX Chicken matrix metalloproteinase-2 (MMP-2) protein sequence.

XX Matrix metalloproteinase-2; MMP-2; alpha-v-beta-5 antagonist; treatment;

XX vitronectin receptor; inhibition; angiogenesis; integrin; tumour growth;

XX restenosis; neovascularisation.

XX Gallus sp.

XX WO9745447-A1.

XX 04-DEC-1997.

XX 30-MAY-1997; 97WO-US09099.

XX 31-MAY-1996; 96US-0018733.

XX 31-MAY-1996; 96US-0015869.

XX (SCRI) SCRIPPS RES INST.

XX Brooks P, Cheresh DA, Friedlander M;

XX WPI; 1998-041758/04.

XX Packaging material containing polypeptide antagonist of alphav,

XX betas integrin - used for inhibition of angiogenesis, and for

XX treating tumours, inflammation, eye diseases etc.

XX Disclosure: Fig 15A-B; 117pp: English.

XX The present sequence represents the chicken matrix metalloproteinase-2

XX (MMP-2) protein sequence. Fragments of this protein (AAW41234-39) are

XX able to act as alpha-v-beta-5 antagonists. Alpha-v-beta-5 is a

XX vitronectin receptor. Inhibitors of alpha-v-beta-5 can inhibit

Best Local Similarity 25.5%; Pred. No. 3.6;
Matches 28; Conservative 9; Mismatches 37; Indels 36; Gaps 5;

```

OY 4 LVVFLLMGVTWGPVTEA-----AIFYETQXSLMAESEHXKTLGGCCDADVP6 51
    | : |||| | | | : || : || | |
DB 12 lltclllwvpvggevvnatkavltlqppwvs1f9kenvclwceghl----- 58
OY 52 PGDSRLPAVQEW---GAQEPVHLDSPAIKHQLTGTQGRYRCRSGLS 98
    |||| | : | : | : ||| : | |
DB 59 -p9ds-----tqwfingtavqistpsyslpe--asfqdsgeyrcq19ss 100

```

Search completed: January 7, 2002, 16:49:15
Job time: 307 sec

•
•
•
•

.

✓


```

: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 233 South Wacker Drive/6300 Sears Tower
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/402,002
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 1997-71082
: FILING DATE: 25-MAR-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP98/01286
: FILING DATE: 23-MAR-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Cawley, Jr., Thomas A.
: REGISTRATION NUMBER: 40,944
: REFERENCE/DOCKET NUMBER: 19036/36276
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 474-6300
: TELEFAX: (312) 474-0448
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1747 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Klebsiella pneumoniae
: STRAIN: Clinical Isolate KP-85-43
: US-09-402-002-2

alignment_scores:
      Quality: 47.50      Length: 18
      Ratio: 3.167      Gaps: 1
      Percent Similarity: 83.333      Percent Identity: 44.444

alignment_block:
US-09-471-276-831_COPY_1_16 x US-09-402-002-2/rev ..
Align seg 1/1 to reverse of: US-09-402-002-2 from: 1 to: 1747

1 MetSerMetLeuValAlaPheLeuLeu.....TrpGlyValTrp 14
: ::::::::::::::::::::| 1442
: 1442 GTGACATATTGATTATTCCTGCTGTCGATGACCTGGGACAC 1393
      14 TTP 15
      1392 ATGG 1389

seq_name: /cgn2_6/plodata/2/1na/6A_COMB.seq:US-09-027-064-3

seq_documentation_block:
: Sequence 3, Application US/09027064
: Patent No. 6133006
: GENERAL INFORMATION:
: APPLICANT: SHABON, USMAN
: APPLICANT: BERGSMAN, DEBK
: TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE
: PROTEIN KINASE-HTLAR33
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
```

```

: ADDRESSEE: RATNER & PRESTIA
: STREET: P.O. BOX 980
: CITY: VALLEY FORGE
: STATE: PA
: COUNTRY: USA
: ZIP: 19482
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/027,064
: FILING DATE: 20-FEB-1998
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/053,924
: FILING DATE: 28-JUL-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: PRESTIA, PAUL F
: REGISTRATION NUMBER: 23,031
: REFERENCE/DOCKET NUMBER: GH-70172
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-407-0700
: TELEFAX: 610-407-0701
: TELEX: 846169
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1338 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-09-027-064-3

alignment_scores:
      Quality: 46.00      Length: 14
      Ratio: 4.182      Gaps: 0
      Percent Similarity: 78.571      Percent Identity: 57.143

alignment_block:
US-09-471-276-831_COPY_1_16 x US-09-027-064-3 ..
Align seg 1/1 to: US-09-027-064-3 from: 1 to: 1338

2 SerMetLeuValAlaPheLeuLeuTrpGlyValTrp 15
: ||| |||:||||| 845
: 845 TCCCGCCTCCTGATGACCCCTGCCACCTCTGGGCGCCACCTGG 886

seq_name: /cgn2_6/plodata/2/1na/6B_COMB.seq:US-09-271-815-3

seq_documentation_block:
: Sequence 3, Application US/09271815
: Patent No. 6297036
: GENERAL INFORMATION:
: APPLICANT: BERGSMAN, DEBK
: APPLICANT: SHABON, USMAN
: TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE PROTEIN KINASE-HTLAR33
: FILE REFERENCE: GH-70172-1
: CURRENT APPLICATION NUMBER: US/09/271,815
: EARLIER FILING DATE: 1999-03-18
: EARLIER APPLICATION NUMBER: 09/027,064
: EARLIER FILING DATE: 1998-02-20
: EARLIER APPLICATION NUMBER: 60/053,924
: EARLIER FILING DATE: 1997-07-28
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 3
: LENGTH: 1338
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-271-815-3
```

alignment_scores:

Quality: 46.00 Length: 14
Ratio: 4.182 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 57.143

alignment_block:

US-09-471-276-831_COPY_1_16 x US-09-271-815-3 ..

Align seg 1/1 to: US-09-271-815-3 from: 1 to: 1338

2 SermetleuVal1PheuleuleuTrpGlyValThrTrp 15
||| |||:||||| ||| |||:|||||:|||||
845 TCCCGCTCTGATGACCTGCGCACTCTGGGCTCACCCTGG 886

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-027-064-1

seq_documentation_block:

Sequence 1, Application US/09027064
Patent No. 6133006
GENERAL INFORMATION:
APPLICANT: SHABON, USMAN
APPLICANT: BERGSMÄ, DERK
TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE
TITLE OF INVENTION: PROTEIN KINASE-HTLAR33
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: RAYNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTED for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/027, 064
FILING DATE: 20-FEB-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/053, 924
FILING DATE: 28-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70172
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2394 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-027-064-1

alignment_scores:

Quality: 46.00 Length: 14
Ratio: 4.182 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 57.143

alignment_block:

US-09-471-276-831_COPY_1_16 x US-09-027-064-1 ..

Align seg 1/1 to: US-09-027-064-1 from: 1 to: 2394

2 SermetleuVal1PheuleuleuTrpGlyValThrTrp 15

||| |||:||||| ||| |||:|||||:|||||
1901 TCCCGCTCTGATGACCTGCGCACTCTGGGCTCACCCTGG 1942

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-271-815-1

seq_documentation_block:

Sequence 1, Application US/09271815
Patent No. 6297036
GENERAL INFORMATION:
APPLICANT: BERGSMÄ, DERK
APPLICANT: SHABON, USMAN
TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE PROTEIN KINASE-HTLAR33
FILE REFERENCE: GH-70172-1
CURRENT APPLICATION NUMBER: US/09/271, 815
CURRENT FILING DATE: 1999-03-18
EARLIER APPLICATION NUMBER: 09/027, 064
EARLIER FILING DATE: 1998-02-20
EARLIER APPLICATION NUMBER: 60/053, 924
EARLIER FILING DATE: 1997-07-28
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2394
TYPE: DNA
ORGANISM: Homo sapiens
US-09-271-815-1

alignment_scores:

Quality: 46.00 Length: 14
Ratio: 4.182 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 57.143

alignment_block:

US-09-471-276-831_COPY_1_16 x US-09-271-815-1 ..

Align seg 1/1 to: US-09-271-815-1 from: 1 to: 2394

2 SermetleuVal1PheuleuleuTrpGlyValThrTrp 15
||| |||:||||| ||| |||:|||||:|||||
1901 TCCCGCTCTGATGACCTGCGCACTCTGGGCTCACCCTGG 1942

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-816-693A-1

seq_documentation_block:

Sequence 1, Application US/08816693A
Patent No. 5874241
GENERAL INFORMATION:
APPLICANT: Takahashi, Joseph S
APPLICANT: Turek, Fred W
APPLICANT: Pinto, Lawrence H
TITLE OF INVENTION: Clock Gene and Gene Product
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dressler, Rocky, Milnamow & Katz
STREET: Two Prudential Plaza, Suite 4700
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/816, 693A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5874241thrnp, Thomas E

```
REGISTRATION NUMBER: 33,268
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-616-5400
TELEFAX: 312-616-5460
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7498 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 389..2954
US-08-816-693A-1
```

```
alignment_scores:
Quality: 46.00 Length: 9
Ratio: 5.111 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889
```

alignment_block:

```
US-09-471-276-831_COPY_1_16 x US-08-816-693A-1 ..
```

```
Align seg 1/1 to: US-08-816-693A-1 from: 1 to: 7498
```

```
8 LeuLeuLeuTrpGlyValThrTrpGly 16
|||||
```

```
5598 CTGCTTCTGTATGGGTGACTTGCGGT 5624
```

```
seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq:US-08-885-291-1
```

seq_documentation_block:

```
Sequence 1, Application US/08885291A
Patent No. 6057125
GENERAL INFORMATION:
APPLICANT: Takahashi, Joseph S.
APPLICANT: Turek, Fred W.
TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
FILE REFERENCE: 0290-5
CURRENT APPLICATION NUMBER: US/08/885,291A
CURRENT FILING DATE: 1997-06-30
EARLIER APPLICATION NUMBER: 08/816,693
EARLIER FILING DATE: 1997-03-13
NUMBER OF SEQ ID NOS: 55
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1
LENGTH: 7498
TYPE: DNA
ORGANISM: Mus musculus
US-08-885-291-1
```

```
alignment_scores:
Quality: 46.00 Length: 9
Ratio: 5.111 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889
```

alignment_block:

```
US-09-471-276-831_COPY_1_16 x US-08-885-291-1 ..
```

```
Align seg 1/1 to: US-08-885-291-1 from: 1 to: 7498
```

```
8 LeuLeuLeuTrpGlyValThrTrpGly 16
|||||
```

```
5598 CTGCTTCTGTATGGGTGACTTGCGGT 5624
```

```
seq_name: /cgn2_6/plodata/2/lna/6B_COMB.seq:US-09-496-672-1
```

seq_documentation_block:

```
Sequence 1, Application US/09496672
```

```
Patent No. 6291429
GENERAL INFORMATION:
APPLICANT: Takahashi, Joseph S.
APPLICANT: Turek, Fred W.
APPLICANT: Pinto, Lawrence H.
TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
FILE REFERENCE: 0290-5
CURRENT APPLICATION NUMBER: US/09/496,672
CURRENT FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: 08/885,291
PRIOR FILING DATE: 1997-06-30
PRIOR APPLICATION NUMBER: 08/816,693
PRIOR FILING DATE: 1997-03-13
NUMBER OF SEQ ID NOS: 55
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1
LENGTH: 7498
TYPE: DNA
ORGANISM: Mus musculus
US-09-496-672-1
```

```
alignment_scores:
Quality: 46.00 Length: 9
Ratio: 5.111 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889
```

alignment_block:

```
US-09-471-276-831_COPY_1_16 x US-09-496-672-1 ..
```

```
Align seg 1/1 to: US-09-496-672-1 from: 1 to: 7498
```

```
8 LeuLeuLeuTrpGlyValThrTrpGly 16
|||||
```

```
5598 CTGCTTCTGTATGGGTGACTTGCGGT 5624
```

```
seq_name: /cgn2_6/plodata/2/lna/6B_COMB.seq:US-09-172-108-48
```

seq_documentation_block:

```
Sequence 48 Application US/09172108
Patent No. 6160104
GENERAL INFORMATION:
APPLICANT: Cunningham, Mary Jane
APPLICANT: Zweigler, Gary B.
APPLICANT: Panzer, Scott R.
APPLICANT: Selhammer, Jeffrey J.
TITLE OF INVENTION: MARKERS FOR PEROXISOMAL PROLIFERATORS
FILE REFERENCE: PA-0012 US
CURRENT APPLICATION NUMBER: US/09/172,108
CURRENT FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 56
SOFTWARE: PERL Program
SEQ ID NO 48
LENGTH: 285
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: 700607235H1
US-09-172-108-48
```

```
alignment_scores:
Quality: 45.00 Length: 16
Ratio: 3.750 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 56.250
```

alignment_block:

```
US-09-471-276-831_COPY_1_16 x US-09-172-108-48 ..
```

```
Align seg 1/1 to: US-09-172-108-48 from: 1 to: 285
```

```
1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
|||||
```



```
14 ATGCTCTGCTGACTGACTGACTGCTGCGGTTTCATTCGTGGGC 61
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-172-711-45
seq_documentation_block:
; Sequence 45, Application US/09172711
; Patent No. 6160105
; GENERAL INFORMATION:
; APPLICANT: Cunningham, Mary Jane
; APPLICANT: Zweiger, Gary B.
; APPLICANT: Panzer, Scott R.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: MONITORING TOXICOLOGICAL RESPONSES
; FILE REFERENCE: PA-0011 US
; CURRENT APPLICATION NUMBER: US/09/172,711
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PERL Program
; SEQ ID NO 45
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 700607235H1
; US-09-172-711-45

alignment_scores:
Quality: 45.00 Length: 16
Ratio: 3.750 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 56.250

alignment_block:
US-09-471-276-831_COPY_1_16 x US-09-172-711-45 ..
Align seg 1/1 to: US-09-172-711-45 from: 1 to: 285

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
|||||:|||||:|||||:|||||:|||||
14 ATGCTCTGCTGACTGACTGACTGCTGCGGTTTCATTCGTGGGC 61

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-254-573-1
seq_documentation_block:
; Sequence 1, Application US/08254573
; Patent No. 5610032
; GENERAL INFORMATION:
; APPLICANT: KAMBOJ, Rajender
; APPLICANT: ELIOTT, Candace
; APPLICANT: NUTT, Stephen
; TITLE OF INVENTION: AMPA-BINDING HUMAN GLUR1 RECEPTORS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,573
; FILING DATE: 06-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/896,611
; FILING DATE: 10-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
```

```
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16777/179 ALLE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3220 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 62..2782
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 62..115
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 116..2782
US-08-254-573-1

alignment_scores:
Quality: 45.00 Length: 16
Ratio: 3.462 Gaps: 0
Percent Similarity: 81.250 Percent Identity: 50.000

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-254-573-1/rev ..
Align seg 1/1 to reverse of: US-08-254-573-1 from: 1 to: 3220

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
:::|||||:|||||:|||||:|||||:|||||
2875 CTTTCATTGTTGTTGTTGTTTGGCAGTGAGGGTTTGAGATGGGCG 2878

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-264-419C-3
seq_documentation_block:
; Sequence 3, Application US/09264419C
; Patent No. 6174682
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: No. 6174682el Thioresdoxin Family Active Site Molecules and us
; FILE REFERENCE: MNI-076
; CURRENT APPLICATION NUMBER: US/09/264,419C
; CURRENT FILING DATE: 1999-03-08
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 840
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(840)
; US-09-264-419C-3

alignment_scores:
Quality: 44.00 Length: 13
Ratio: 4.400 Gaps: 0
Percent Similarity: 76.923 Percent Identity: 53.846

alignment_block:
US-09-471-276-831_COPY_1_16 x US-09-264-419C-3 ..
Align seg 1/1 to: US-09-264-419C-3 from: 1 to: 840
```

```
3 MetLeuValValPheLeuLeuLeuTrpGlyValThrTrp 15
::: |||::: ||| ||| ||| ||| ||| ||| ||| ||| |||
31 CTGCACGCTCCTGCTGCTGTTGCTTGGGCTGCTCCCTGG 69

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-264-419C-1

seq_documentation_block:
: Sequence 1, Application US/09264419C
: Patent No. 6174682
: GENERAL INFORMATION:
: APPLICANT: Khodadoust, Mehran
: TITLE OF INVENTION: No. 6174682el Thiorodoxin Family Active Site Molecules and Uses
: FILE REFERENCE: MNI-076
: CURRENT APPLICATION NUMBER: US/09/264,419C
: CURRENT FILING DATE: 1999-03-08
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 1
: LENGTH: 1207
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (74)..(913)
: OTHER INFORMATION: AT POSITION 1050 N.ANY NUCLEIC ACID
US-09-264-419C-1

Alignment_scores:
Quality: 44.00 Length: 13
Ratio: 4.400 Gaps: 0
Percent Similarity: 76.923 Percent Identity: 53.846

alignment_block:
US-09-471-276-831_COPY_1_16 x US-09-264-419C-1 ..

Align seg 1/1 to: US-09-264-419C-1 from: 1 to: 1207

3 MetLeuValValPheLeuLeuLeuTrpGlyValThrTrp 15
::: |||::: ||| ||| ||| ||| ||| ||| ||| ||| |||
104 CTGCACGCTCCTGCTGCTGTTGCTTGGGCTGCTCCCTGG 142

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-033-428-1

seq_documentation_block:
: Sequence 1, Application US/09033428
: Patent No. 6254862
: GENERAL INFORMATION:
: APPLICANT: Little, Andrew
: APPLICANT: Lamparski, Henry
: APPLICANT: Schuur, Eric
: APPLICANT: Henderson, Daniel
: TITLE OF INVENTION: ADENOVIRUS VECTORS SPECIFIC FOR CELLS
: TITLE OF INVENTION: EXPRESSING APHA-FETOPROTEIN AND METHODS OF USE THEREOF
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 755 PAGE MILL ROAD
: CITY: PALO ALTO
: STATE: CA
: COUNTRY: USA
: ZIP: 94304-1018
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/033,428
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
```

```
: NAME: POLIZZI, CATHERINE M.
: REGISTRATION NUMBER: 40,130
: REFERENCE/DOCKET NUMBER: 34802-30004.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 813-5600
: TELEFAX: (415) 494-0792
: TELE: 706141 MRSNPOERS SFO
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 822 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-09-033-428-1

Alignment_scores:
Quality: 43.00 Length: 15
Ratio: 3.071 Gaps: 0
Percent Similarity: 93.333 Percent Identity: 46.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-09-033-428-1/rev ..

Align seg 1/1 to reverse of: US-09-033-428-1 from: 1 to: 822

2 SerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGly 16
::: |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
737 AACCTTTGTTGTATATCTGTACTTATGATTTCTTACGT 693

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-931-608A-1

seq_documentation_block:
: Sequence 1, Application US/08931608A
: Patent No. 6302685
: GENERAL INFORMATION:
: APPLICANT: Lobel, Peter
: APPLICANT: Sleat, David E.
: TITLE OF INVENTION: NOVEL HUMAN LYSOSOMAL PROTEIN AND METHODS OF ITS USE
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David A. Jackson, Esq.
: STREET: 411 Hackensack Ave, Continental Plaza, 4th floor
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/931,608A
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson Esq., David A.
: REGISTRATION NUMBER: 26,742
: REFERENCE/DOCKET NUMBER: 601-1-077
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-487-5800
: TELEFAX: 201-343-1684
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3487 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
US-08-931-608A-1
```



```

STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DO5
SOFTWARE: PatentIn Release #1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/981,459
FILING DATE: 19-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: HOLMAN, John C.
REGISTRATION NUMBER: 22,769
REFERENCE/DOCKET NUMBER: P61784US0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-638-6666
TELEFAX: 202-393-5350
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12687 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
OS-08-981-459-1

```

```
alignment_scores:
  Quality: 43.00
  Ratio: 3.583
  Percent Similarity: 100.000
  Length: 12
  Gaps: 0
  Percent Identity: 58.333
```

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-981-459-1 .

Align seg 1/1 to: US-08-981-459-1 from: 1 to: 12687

```

1 MetSerMetLeuValValPheLculeLeuLeuTrpGly 12
:::|||||:::|||||:::|||||
11132 TTATCTATGATTCGATTCGCTATTCCTGCTTTGGGGT 11167

```

```
seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-328-111-517
seq_documentation_block:
```

```

? Patent NO. 6262333
? GENERAL INFORMATION:
? APPLICANT: Endege, Wilson O.
? APPLICANT: Steinmann, Kathleen E.
? APPLICANT: Aslie, Jon H.
? APPLICANT: Burgess, Christopher C.
? APPLICANT: Bushnell, Steven E.
? APPLICANT: Carroll III, Eddie
? APPLICANT: Catino, Theodore J.
? APPLICANT: Dettl, Adrian
? APPLICANT: Ford, Donna M.
? APPLICANT: Lewis, Marcia E.
? APPLICANT: Monahan, John E.
? APPLICANT: Schlegel, Robert
? TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
? FILE REFERENCE: CCD-257 (US)
? CURRENT APPLICATION NUMBER: US/09/328,111
? CURRENT FILING DATE: 1999-06-08
? EARLIER APPLICATION NUMBER: US 60/088,801
? EARLIER FILING DATE: 1998-06-10
? NUMBER OF SEQ. ID NOS: 850
? SOFTWARE: FASTSEQ for Windows Version 3.0
? SEQ ID NO 517
?
? LENGTH: 611
? TYPE: DNA

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(611)
; OTHER INFORMATION: n " A,T,C or G
US-09-328-111-517

```

alignment_scores:	
Quality:	42.50
Ratio:	3.542
Percent Similarity:	80.000
	Length: 15
	Gaps: 1
Percent Identity:	66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-09-328-111-517 .

Align seg 1/1 to: US-09-328-111-517 from: 1 to: 613

```

3 MetLeuValValIheLeuLeuLeuTrp..GlyValThrTrrGly 16
|||||  |||||  |||  :::::|||||
361 ATGCTCGTGGCTTCCTCTTGGACATGGGTCTCAACACCACTGGGGA 405

```

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-583-318-4

```
seq_documentation_block:
; Sequence 4, Application US/08583318
; Patent NO. 5693483
SEQUENCE INFORMATION
```

APPLICANT: Staunton, Donald

TITLE OF INVENTION: Cytoplasmic Modulators of Integrin

NUMBER OF SEQUENCES: 14

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

CITY: Chicago

COUNTRY: USA

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25
;
***** ADDITIONAL DATA *****

APPLICATION NUMBER: US/08/583,318

CLASSIFICATION: 435

NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38 650

REFERENCE/DOCKET NUMBER: 27866/33050

TELEPHONE: 312-474-6300

INFORMATION FOR SEQ ID NO: 4:

LENGTH: 1631 base pairs

STRANDEDNESS: single

MOLECULE TYPE: CDNA

	NAME/KEY:	CDS
;		
LOCATION		

US-08-583-318-4

alignment_scores:		
Quality:	42.50	Length: 15
Ratio:	3.542	Gaps: 1
Percent Similarity:	80.000	Percent Identity: 66.667

alignment_block:

US-09-471-276-831_COPY_1_16 x US-08-583-318-4 ..

Align seg 1/1 to: US-08-583-318-4 from: 1 to: 1631

3 MetLeuValValPheLeuLeuLeuTrp...GlyValThrTrpGly 16
||||||| ||||||| ||| :::::|||||||
374 ATGCTGCTTGCTCTCTCTTGAGACTGGGCTCAACACCTGGGGA 418

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-699-103B-9

seq_documentation_block:

; Sequence 9, Application US/08699103B

; Patent No. 6107462

; GENERAL INFORMATION:

; APPLICANT: Rine, Jasper D.

; APPLICANT: Hampton, Randolph

; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING

; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 2200 Sand Hill Road, Suite 100

; CITY: Menlo Park

; STATE: CA

; COUNTRY: USA

; ZIP: 94025

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/699,103B

; FILING DATE: 16-AUG-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/002,581

; FILING DATE: 17-AUG-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Green, Grant D.

; REGISTRATION NUMBER: 31,259

; REFERENCE/DOCKET NUMBER: 09272/005001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650/322-5070

; TELEFAX: 650/854-0875

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2556 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: Genomic DNA

US-08-699-103B-9

alignment_scores:

Quality: 42.50 Length: 15

Ratio: 3.542 Gaps: 1

Percent Similarity: 80.000 Percent Identity: 66.667

alignment_block:

US-09-471-276-831_COPY_1_16 x US-08-699-103B-9 ..

Align seg 1/1 to: US-08-699-103B-9 from: 1 to: 2556

3 MetLeuValValPheLeuLeuLeuTrp...GlyValThrTrpGly 16
||||||| ||||||| ||| :::::|||||||
1599 ATGCTGCTTGCTCTCTCTTGAGACTGGGCTCAACACCTGGGGA 1643

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-699-103B-11

seq_documentation_block:

; Sequence 11, Application US/08699103B

; Patent No. 6107462

; GENERAL INFORMATION:

; APPLICANT: Rine, Jasper D.

; APPLICANT: Hampton, Randolph

; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING

; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 2200 Sand Hill Road, Suite 100

; CITY: Menlo Park

; STATE: CA

; COUNTRY: USA

; ZIP: 94025

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/699,103B

; FILING DATE: 16-AUG-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/002,581

; FILING DATE: 17-AUG-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Green, Grant D.

; REGISTRATION NUMBER: 31,259

; REFERENCE/DOCKET NUMBER: 09272/005001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650/322-5070

; TELEFAX: 650/854-0875

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2731 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: Genomic DNA

US-08-699-103B-11

alignment_scores:

Quality: 42.50 Length: 15

Ratio: 3.542 Gaps: 1

Percent Similarity: 80.000 Percent Identity: 66.667

alignment_block:

US-09-471-276-831_COPY_1_16 x US-08-699-103B-11 ..

Align seg 1/1 to: US-08-699-103B-11 from: 1 to: 2731

3 MetLeuValValPheLeuLeuLeuTrp...GlyValThrTrpGly 16
||||||| ||||||| ||| :::::|||||||
1694 ATGCTGCTTGCTCTCTCTTGAGACTGGGCTCAACACCTGGGGA 1738

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-972-927-1

seq_documentation_block:

; Sequence 1, Application US/08972927

; Patent No. 6166290

; GENERAL INFORMATION:

; APPLICANT: Rea, Phillip A

; APPLICANT: Lu, Yu-ping

; APPLICANT: Li, Ze-sheng

; TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN

; TITLE OF INVENTION: PLANTS

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.

; STREET: One Commerce Square, 2005 Market Street, 22nd

; STREET: Floor

CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: US
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,927
FILING DATE: 18-NOV-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,040
FILING DATE: 18-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/061,328
FILING DATE: 08-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Doyle Leary Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9596-1202
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-972-927-1

alignment_scores:
Quality: 42.50 Length: 15
Ratio: 3.542 Gaps: 1
Percent Similarity: 80.000 Percent Identity: 53.333

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-972-927-1 ..

Align seg 1/1 to: US-08-972-927-1 from: 1 to: 5232

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrp 15
||||| |||::: :::::|||||::: |||
564 ATGTGAGCTTCTGCTCATATATGCTCTTGTGGCGGATA...TGG 605

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-972-927-2

seq_documentation_block:
Sequence 2, Application US/08972927
Patent No. 6166290
GENERAL INFORMATION:
APPLICANT: Rea, Phillip A
APPLICANT: Lu, Yu-Ping
APPLICANT: Li, Ze-Sheng
TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: One Commerce Square, 2005 Market Street, 22nd
STREET: Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: US
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,927
FILING DATE: 18-NOV-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,040
FILING DATE: 18-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/061,328
FILING DATE: 08-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Doyle Leary Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9596-1202
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 9936 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-972-927-2

alignment_scores:
Quality: 42.50 Length: 15
Ratio: 3.542 Gaps: 1
Percent Similarity: 80.000 Percent Identity: 53.333

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-972-927-2 ..

Align seg 1/1 to: US-08-972-927-2 from: 1 to: 9936

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrp 15
||||| |||::: :::::|||||::: |||
951 ATGTGAGCTTCTGCTCATATATGCTCTTGTGGCGGATA...TGG 992

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-276-531-5

seq_documentation_block:
Sequence 5, Application US/09276531
Patent No. 6183968
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Reddy, Roopa
APPLICANT: Guegler, Karl J.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
STREET: PALO ALTO
STATE: CALIFORNIA
CITY: PALO ALTO
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/276.531
: FILING DATE: Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/079,677
: FILING DATE: March 27, 1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Lynn E. Murry, Ph.D.
: REGISTRATION NUMBER: 42,918
: REFERENCE/DOCKET NUMBER: PA-0008 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 845-0555
: TELEFAX: (650) 845-4166
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1350 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: PANCYT01
: CLONE: 1514169
: US-09-276-531-5

alignment_scores:
      Quality: 42.00      Length: 9
      Ratio: 5.250      Gaps: 0
      Percent Similarity: 88.889      Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-09-276-531-5 ..
Align seq 1/1 to: US-09-276-531-5 from: 1 to: 1350
      8 LeuLeuLeuTrrpGlyValThrtTrpGly 16
      |||||::|||:::|||||
      726 TTGCTGATGTGCTACATCGGGG 752

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-632-470-53

seq_documentation_block:
: Sequence 53, Application US/08632470
: Patent No. 5976791
: GENERAL INFORMATION:
: APPLICANT: MABILAT, CLAUDE
: TITLE OF INVENTION: NUCLEOTIDE FRAGMENTS CAPABLE OF
: TITLE OF INVENTION: HYBRIDIZING SPECIFICALLY TO RICKETTSIA RDNA OR RRNA AND
: TITLE OF INVENTION: THEIR USE AS PROBES OR PRIMERS
: NUMBER OF SEQUENCES: 53
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OLIEF & BERRIDGE
: STREET: P O BOX 19928
: CITY: ALEXANDRIA
: STATE: VA
: COUNTRY: USA
: ZIP: 22320
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/632,470
: FILING DATE: 08-JUL-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: BERRIDGE, WILLIAM P
: REGISTRATION NUMBER: 30,024
: REFERENCE/DOCKET NUMBER: WPB 38238
```

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)-836-6400
: TELEFAX: (703)-836-2787
: INFORMATION FOR SEQ ID NO: 53:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1484 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-632-470-53

alignment_scores:
      Quality: 42.00      Length: 6
      Ratio: 7.000      Gaps: 0
      Percent Similarity: 100.000      Percent Identity: 83.333

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-632-470-53 ..
Align seq 1/1 to: US-08-632-470-53 from: 1 to: 1484
      11 TrpGlyValThrtTrpGly 16
      |||||:::|||||
      143 TGGGGCATTAACCTGGGGA 160

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-299-810A-27

seq_documentation_block:
: Sequence 27, Application US/08299810A
: Patent No. 5721097
: GENERAL INFORMATION:
: APPLICANT: Rossau, Rudi
: APPLICANT: Van Heuverswyn, Hugo
: TITLE OF INVENTION: HYBRIDIZATION PROBES FOR THE
: TITLE OF INVENTION: DETECTION OF BRANHAMELLA CATARRHALIS STRAINS
: NUMBER OF SEQUENCES: 28
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merchant & Gould
: STREET: 3100 No. 5721097west Center
: CITY: Minneapolis
: STATE: MN
: COUNTRY: USA
: ZIP: 55402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/299,810A
: FILING DATE: 01-SEP-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Hillson, Randall A.
: REGISTRATION NUMBER: 31,838
: REFERENCE/DOCKET NUMBER: 8076.70-US-WO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-332-5300
: TELEFAX: 612-332-9081
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1485 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Branhamella catarrhalis
: IMMEDIATE SOURCE:
: CLONE: 16S rRNA Gene
: US-08-299-810A-27
```

```
alignment_scores:          Quality: 42.00          Length: 6
                          Ratio: 7.000          Gaps: 0
                          Percent Similarity: 100.000          Percent Identity: 83.333

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-299-810A-27  ..
Align seg 1/1  to: US-08-299-810A-27  from: 1  to: 1485

11 TrpGlyValThrTrpGly 16
|||||:|||||
95 TCGCGCATACTTGGCGA 112

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-864-224-1

seq_documentation_block:
; Sequence 1, Application US/08864224
; Patent No. 5851808
; GENERAL INFORMATION:
; APPLICANT: Elledge, Stephen J.
; APPLICANT: Liu, Qinghua
; TITLE OF INVENTION: Rapid subcloning Using Site-Specific
; TITLE OF INVENTION: Recombination
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,224
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: BCH-02681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc .. "DNA"
US-08-864-224-1

alignment_scores:          Quality: 42.00          Length: 9
                          Ratio: 5.250          Gaps: 0
                          Percent Similarity: 88.889          Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-864-224-1  ..
Align seg 1/1  to: US-08-864-224-1  from: 1  to: 2220

8 LeuLeuLeuTrpGlyValThrTrpGly 16
:::|||||:|||||  |||||
```

```
673 ATTCTATTCTGGGGGTGGGGTGGCGC 699

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-801-092-5

seq_documentation_block:
; Sequence 5, Application US/08801092
; Patent No. 6074850
; GENERAL INFORMATION:
; APPLICANT: Antelman, Douglas
; APPLICANT: Gregory, Richard J.
; APPLICANT: Mills, Kenneth N.
; TITLE OF INVENTION: Tissue Specific Expression of
; TITLE OF INVENTION: Retinoblastoma Protein
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,092
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/751,517
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 016930-001020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 703-576-0300
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3853 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 209..250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 254..289
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 293..505
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 509..514
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 518..520
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 524..658
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 662..691
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 695..748
; FEATURE:
```



```
;
; NAME/KEY: CDS
; LOCATION: 752..781
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 785..829
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1132..1134
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1138..1149
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 833..862
; US-08-801-092-5

alignment_scores:
    Quality: 42.00      Length: 9
    Ratio: 5.250        Gaps: 0
    Percent Similarity: 88.889    Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-801-092-5 ..
Align seg 1/1 to: US-08-801-092-5 from: 1 to: 3853

      8 LeuleuleuTrpGlyValThrTrpGly 16
      :::::::::::||||| 11111
      1307 ATTCTATTCTGGGGGTGGGGTGGGC 1333

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-801-092-19
seq_documentation_block:
; Sequence 19, Application US/08801092
; Patent No. 6074850
; GENERAL INFORMATION:
; APPLICANT: Antelman, Douglas
; APPLICANT: Gregory, Richard J.
; APPLICANT: Wils, Kenneth N.
; TITLE OF INVENTION: Tissue Specific Expression of
; TITLE OF INVENTION: Retinoblastoma Protein
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,092
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/751,517
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 016930-001020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 703-576-0300
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4026 base pairs
```

```
;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 209..250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 234..289
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 293..505
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 509..514
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 518..520
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 524..658
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 662..691
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 695..748
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 752..781
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 785..829
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 833..862
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1305..1307
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1311..1322
; US-08-801-092-19

alignment_scores:
    Quality: 42.00      Length: 9
    Ratio: 5.250        Gaps: 0
    Percent Similarity: 88.889    Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-801-092-19 ..
Align seg 1/1 to: US-08-801-092-19 from: 1 to: 4026

      8 LeuleuleuTrpGlyValThrTrpGly 16
      :::::::::::||||| 11111
      1480 ATTCTATTCTGGGGGTGGGGTGGGC 1506

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-801-092-33
seq_documentation_block:
; Sequence 33, Application US/08801092
; Patent No. 6074850
; GENERAL INFORMATION:
; APPLICANT: Antelman, Douglas
; APPLICANT: Gregory, Richard J.
; APPLICANT: Wils, Kenneth N.
; TITLE OF INVENTION: Tissue Specific Expression of
; TITLE OF INVENTION: Retinoblastoma Protein
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/801,092
FILING DATE: 14-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/751,517
FILING DATE: 15-NOV-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Flits, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 016930-001020
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 703-576-0300
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 4249 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 209..250
FEATURE:
NAME/KEY: CDS
LOCATION: 254..289
FEATURE:
NAME/KEY: CDS
LOCATION: 293..505
FEATURE:
NAME/KEY: CDS
LOCATION: 509..514
FEATURE:
NAME/KEY: CDS
LOCATION: 518..520
FEATURE:
NAME/KEY: CDS
LOCATION: 524..658
FEATURE:
NAME/KEY: CDS
LOCATION: 662..691
FEATURE:
NAME/KEY: CDS
LOCATION: 695..748
FEATURE:
NAME/KEY: CDS
LOCATION: 752..781
FEATURE:
NAME/KEY: CDS
LOCATION: 785..829
FEATURE:
NAME/KEY: CDS
LOCATION: 833..862
FEATURE:
NAME/KEY: CDS
LOCATION: 1528..1530
FEATURE:
NAME/KEY: CDS
LOCATION: 1534..1545
US-08-801-092-33

alignment_scores:
Quality: 42.00 Length: 9
Ratio: 5.250 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 66.667
alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-801-092-33 ..
Align seg 1/1 to: US-08-801-092-33 from: 1 to: 4249
8 LeuLeuLeuTrpGlyValThrTrpGly 16
:::|||||:||||| 11111
1703 ATTCTATTCTGGGGGCTGGGGTGGGC 1729
seq_name: /cgn2_6/plodata/2/1na/5A_COMB.seq:US-08-343-401A-3
seq_documentation_block:
Sequence 3, Application US/08343401A
Patent No. 5661132
GENERAL INFORMATION:
APPLICANT: Swain, William F.
APPLICANT: Macklin, Michael D.
APPLICANT: Eriksson, Elof
APPLICANT: Andree, Christophe
TITLE OF INVENTION: Improved Wound Healing
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: PO Box 2113
CITY: Madison
STATE: WI
COUNTRY: USA
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,401A
FILING DATE: 22-NOV-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 11-229-9103-9
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
FAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4283 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: pWRG1630
FEATURE:
NAME/KEY: exon
LOCATION: 713..721
FEATURE:
NAME/KEY: exon
LOCATION: 981..1253
FEATURE:
NAME/KEY: CDS
LOCATION: join(713..721, 981..1253)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 713..1049
US-08-343-401A-3

```
alignment_scores:
  Quality: 42.00      Length: 9
  Ratio: 5.250        Gaps: 0
  Percent Similarity: 88.889   Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-343-401A-3 ..
Align seg 1/1 to: US-08-343-401A-3 from: 1 to: 4283

      8 LeuLeuLeuTrpGlyValThrTrpGly 16
      1480 ATTCTATCTCTGGGGGTGGGGTGGGC 1506

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-445-265A-1

seq_documentation_block:
: Sequence 1, Application US/08445265A
: Patent No. 5697901
: GENERAL INFORMATION:
: APPLICANT: Eriksson, Elof
: TITLE OF INVENTION: GENE DELIVERY BY MICRONEEDLE INJECTION
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Quarles & Brady
: STREET: 1 South Pinckney Street
: CITY: Madison
: STATE: WI
: COUNTRY: US
: ZIP: 53703
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/445,265A
: FILING DATE:
: CLASSIFICATION: 604
: ATTORNEY/AGENT INFORMATION:
: NAME: Seay, Nicholas J
: REGISTRATION NUMBER: 27386
: REFERENCE/DOCKET NUMBER: 110229,91080
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 608-251-5000
: TELEFAX: 608-251-9166
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4283 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "Plasmid DNA"
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(713..721, 981..1250)
: US-08-445-265A-1

alignment_scores:
  Quality: 42.00      Length: 9
  Ratio: 5.250        Gaps: 0
  Percent Similarity: 88.889   Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-445-265A-1 ..
Align seg 1/1 to: US-08-445-265A-1 from: 1 to: 4283

      8 LeuLeuLeuTrpGlyValThrTrpGly 16
```

```
      1480 ATTCTATCTCTGGGGGTGGGGTGGGC 1506
      :::::::::::|||||
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-990-442-1

seq_documentation_block:
: Sequence 1, Application US/08990442
: Patent No. 6090790
: GENERAL INFORMATION:
: APPLICANT: Eriksson, Elof
: TITLE OF INVENTION: GENE DELIVERY BY MICRONEEDLE INJECTION
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Quarles & Brady
: STREET: 1 South Pinckney Street
: CITY: Madison
: STATE: WI
: COUNTRY: US
: ZIP: 53703
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/990,442
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Berson, Bennett J
: REGISTRATION NUMBER: 37094
: REFERENCE/DOCKET NUMBER: 310558,90028
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 608-251-5000
: TELEFAX: 608-251-9166
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4283 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "Plasmid DNA"
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(713..721, 981..1250)
: US-08-990-442-1

alignment_scores:
  Quality: 42.00      Length: 9
  Ratio: 5.250        Gaps: 0
  Percent Similarity: 88.889   Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-990-442-1 ..
Align seg 1/1 to: US-08-990-442-1 from: 1 to: 4283

      8 LeuLeuLeuTrpGlyValThrTrpGly 16
      1480 ATTCTATCTCTGGGGGTGGGGTGGGC 1506

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-760-615-7

seq_documentation_block:
: Sequence 7, Application US/08760615
: Patent No. 6200959
: GENERAL INFORMATION:
: APPLICANT: Haynes, Joel R
: APPLICANT: Schmaljohn, Connie S
: APPLICANT: Fuller, Deborah L
: APPLICANT: Schmaljohn, Alan
```

APPLICANT: Jahrling, Peter B
TITLE OF INVENTION: GENETIC INDUCTION OF ANTI-VIRAL IMMUNE
TITLE OF INVENTION: RESPONSE AND GENETIC VACCINE FOR FILOVIRUS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/760.615
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Berson, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 110229.91241
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4326 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Expression vector
IMMEDIATE SOURCE:
CLONE: pMRG7077
FEATURE:
NAME/KEY: promoter
LOCATION: 1250..2062
FEATURE:
NAME/KEY: Intron
LOCATION: 2063..2887
OTHER INFORMATION: /function= "Human Cytomegalovirus
FEATURE:
NAME/KEY: polyA_site
LOCATION: 2912..3314
FEATURE:
NAME/KEY: CDS
LOCATION: complement (299..1114)
US-08-760-615-7

alignment_scores:
Quality: 42.00 length: 9
Ratio: 5.250 caps: 0
Percent Similarity: 88.889 Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-760-615-7 ..

Align seg 1/1 to: US-08-760-615-7 from: 1 to: 4326

8 LeuLeuLeuTrpGlyValThrTrpGly 16
:::||||:||||||| 111111
3175 ATTCTATTCTGCGCGGTGGCGTGGCG 3201

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-132-808-1
seq_documentation_block:
; Sequence 1, Application US/09132808

Patent No. 6197332
GENERAL INFORMATION:
APPLICANT: Ronald Zuckermann et al.
TITLE OF INVENTION: Lipid-Conjugated Polyamide Compounds and Related
TITLE OF INVENTION: Compositions and Methods Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/132.808
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fujita, Sharon M.
REGISTRATION NUMBER: 38,459
REFERENCE/DOCKET NUMBER: 1387.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-2706
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4328 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-132-808-1

alignment_scores:
Quality: 42.00 length: 9
Ratio: 5.250 caps: 0
Percent Similarity: 88.889 Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-09-132-808-1 ..

Align seg 1/1 to: US-09-132-808-1 from: 1 to: 4328

8 LeuLeuLeuTrpGlyValThrTrpGly 16
:::||||:||||||| 111111
1862 ATTCTATTCTGCGCGGTGGCGTGGCG 1888

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-910-647-2
seq_documentation_block:
; Sequence 2, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-910-647-2

seq_name: /cgn2_6/plodata/2/ina/5b_COMB.seq:US-08-910-647-4

alignment_scores:
    Quality: 42.00      Length: 9
    Ratio: 5.250        Gaps: 0
    Percent Similarity: 88.889    Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-910-647-2  ..
Align seg 1/1 to: US-08-910-647-2 from: 1 to: 4328

      8 LeuLeuLeuTrpGlyValThrTrpGly 16
      ::::::::::::::::::::
1862 ATTCTATTCTGGGGGTGGGGTGGGCGC 1888

seq_documentation_block:
; Sequence 4, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4818 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-910-647-4
```

```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-910-647-4

alignment_scores:
    Quality: 42.00      Length: 9
    Ratio: 5.250        Gaps: 0
    Percent Similarity: 88.889    Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-910-647-4  ..
Align seg 1/1 to: US-08-910-647-4 from: 1 to: 4818

      8 LeuLeuLeuTrpGlyValThrTrpGly 16
      ::::::::::::::::::::
2354 ATTCTATTCTGGGGGTGGGGTGGGCGC 2380

seq_name: /cgn2_6/plodata/2/ina/5a_COMB.seq:US-08-945-913-1

seq_documentation_block:
; Sequence 1, Application US/08345913
; Patent No. 5641665
; GENERAL INFORMATION:
; APPLICANT: Hobart, Peter
; APPLICANT: Parker, Suzanne
; APPLICANT: Margalith, Michael
; APPLICANT: Khatib, Shirin
; TITLE OF INVENTION: PLASMIDS SUITABLE FOR IL-2 EXPRESSION
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/345,913
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ways Vensko, Nancy
; REGISTRATION NUMBER: 36,298
; REFERENCE/DOCKET NUMBER: VTCL.043A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; TELE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
```

```
; LOCATION: 1689...2159
; OTHER INFORMATION:
US-08-345-913-1
```

```
alignment_scores:
    Quality: 42.00      Length: 9
    Ratio: 5.250        Gaps: 0
    Percent Similarity: 88.889    Percent Identity: 66.667
```

```
alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-345-913-1 ..
```

```
Align seg 1/1 to: US-08-345-913-1 from: 1 to: 4928
```

```
8 LeuLeuLeuTrpGlyValThrTrpGly 16
:::||||:||||| 111111
```

```
2388 ATTCTATTCTGGGGCTGGGCTGGCGC 2414
```

```
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-818-562-1
```

```
seq_documentation_block:
```

```
; Sequence 1, Application US/08818562
; Patent No. 6147055
; GENERAL INFORMATION:
; APPLICANT: Hobart, Peter M.
; APPLICANT: Margalith, Michal
; APPLICANT: Parker, Suzanne E.
; APPLICANT: Khatib, Shirin
; TITLE OF INVENTION: Plasmids suitable for IL-2 Expression
; FILE REFERENCE: 1530.0080001
; CURRENT APPLICATION NUMBER: US/08/818,562
; CURRENT FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: US 08/345,913
; EARLIER FILING DATE: 1994-11-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4928
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1689)..(2159)
US-08-818-562-1
```

```
alignment_scores:
    Quality: 42.00      Length: 9
    Ratio: 5.250        Gaps: 0
    Percent Similarity: 88.889    Percent Identity: 66.667
```

```
alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-818-562-1 ..
```

```
Align seg 1/1 to: US-08-818-562-1 from: 1 to: 4928
```

```
8 LeuLeuLeuTrpGlyValThrTrpGly 16
:::||||:||||| 111111
```

```
2388 ATTCTATTCTGGGGCTGGGCTGGCGC 2414
```

```
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-564-313-1
```

```
seq_documentation_block:
```

```
; Sequence 1, Application US/08564313
; Patent No. 5910488
; GENERAL INFORMATION:
; APPLICANT: Nabel, Elizabeth
; APPLICANT: Nabel, Gary
; APPLICANT: Lew, Denise
; APPLICANT: Marguet, Magda
; TITLE OF INVENTION: PLASMIDS SUITABLE FOR GENE THERAPY
; NUMBER OF SEQUENCES: 2
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knodbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
```

```
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/564,313
; FILING DATE: 01-DEC-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: 08/074,344
; FILING DATE: 07-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: VICAL 033CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; TELEX:
```

```
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4965 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; MOLECULE TYPE: circular
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; IMMEDIATE SOURCE:
; CLONE: HLA-B7 and Beta-2
US-08-564-313-1
```

```
alignment_scores:
    Quality: 42.00      Length: 9
    Ratio: 5.250        Gaps: 0
    Percent Similarity: 88.889    Percent Identity: 66.667
```

```
alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-564-313-1 ..
```

```
Align seg 1/1 to: US-08-564-313-1 from: 1 to: 4965
```

```
8 LeuLeuLeuTrpGlyValThrTrpGly 16
:::||||:||||| 111111
```

```
3021 ATTCTATTCTGGGGCTGGGCTGGCGC 3047
```

```
seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-06069-1
```

```
seq_documentation_block:
```

```
; Sequence 1, Application PC/TUS9406069
; GENERAL INFORMATION:
; APPLICANT: Vical Incorporated
; APPLICANT: Regents of the University of Michigan
; APPLICANT: Nabel, Elizabeth
; APPLICANT: Nabel, Gary
; APPLICANT: Lew, Denise
; APPLICANT: Marguet, Magda
; TITLE OF INVENTION: PLASMIDS SUITABLE FOR GENE THERAPY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knodbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
```

STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06069
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/074,344
FILING DATE: 07-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: VICAL 033VPC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-235-8550
TELEFAX: 619-235-0176
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4965 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: circular
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
STRAIN: HLA-B7 and Beta-2
PCT-US94-06069-1

alignment_scores:
Quality: 42.00 Length: 9
Ratio: 5.250 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x PCT-US94-06069-1 ..
Align seg 1/1 to: PCT-US94-06069-1 from: 1 to: 4965
8 LeuLeuLeuTtPcLyAlThrTtPcLy 16
:::|||||:||||| 111111
3021 ATTCTATTCTGGGGGCGGCGTGGCGC 3047

seq_name: /cgn2_6/plodata/2/lna/6B_COMB.seq:US-08-910-647-3

seq_documentation_block:
Sequence 3, Application US/08910647
Patent No. 6251433
GENERAL INFORMATION:
APPLICANT: Zuckermann et al.
TITLE OF INVENTION: Compositions and Methods for
TITLE OF INVENTION: Polynucleotide Delivery
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,647
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fujita, Sharon M.
REGISTRATION NUMBER: 38,459
REFERENCE/DOCKET NUMBER: 1218,002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-2706
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5107 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: linear
US-08-910-647-3

alignment_scores:
Quality: 42.00 Length: 9
Ratio: 5.250 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 66.667

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-910-647-3 ..
Align seg 1/1 to: US-08-910-647-3 from: 1 to: 5107
8 LeuLeuLeuTtPcLyAlThrTtPcLy 16
:::|||||:||||| 111111
2643 ATTCTATTCTGGGGGCGGCGTGGCGC 2669

seq_name: /cgn2_6/plodata/2/lna/5A_COMB.seq:US-08-073-836-3

seq_documentation_block:
Sequence 3, Application US/08073836
Patent No. 5650306
GENERAL INFORMATION:
APPLICANT: Nabel, Gary J.
APPLICANT: Yang, Zhi-yong
APPLICANT: Liu, Jinsong
TITLE OF INVENTION: RECOMBINANT NUCLEIC ACIDS FOR INHIBITING
TITLE OF INVENTION: HIV GENE EXPRESSION
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States of America
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/073,836
FILING DATE: 07-JUN-1993
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UM 9646
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 5653 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-073-836-3

```

```

alignment_scores:
  Quality: 42.00      Length: 9
  Ratio: 5.250        Gaps: 0
  Percent Similarity: 88.889  Percent Identity: 66.667

```

```

alignment_block:
US-09-471-276-831_COPY_1_16 x US-08-073-836-3  ..

```

```

Align seg 1/1  to: US-08-073-836-3  from: 1  to: 5653

```

```

      8 leuLeuLeuTrpGlyValThrTrpGly 16
      :::::::::::|||||  |||||
1286 ATTCATATCTCGGGCGGCTCGCGCGC 1312

```

```

seq_name: /cgn2_6/plodata/2/1na/5A_COMB.seq:US-08-235-277-1

```

```

seq_documentation_block:

```

```

; Sequence 1, Application US/08235277
; Patent No. 5733543

```

```

; GENERAL INFORMATION:

```

```

; APPLICANT: NABEL, GARY J
; APPLICANT: WOFENDIN, CLIVE

```

```

; APPLICANT: YANG, MIN-SUN

```

```

; APPLICANT: SHEEHY, MICHAEL J

```

```

; TITLE OF INVENTION: INTRODUCTION OF HIV-PROTECTIVE GENES
; TITLE OF INVENTION: INTO CELLS BY PARTICLE-MEDIATED GENE TRANSFER

```

```

; NUMBER OF SEQUENCES: 1

```

```

; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

```

```

; ADDRESSEE: P.C.

```

```

; STREET: 1755 S. Jefferson Davis Highway, Suite 400

```

```

; CITY: Arlington

```

```

; STATE: Virginia

```

```

; COUNTRY: U.S.A.

```

```

; ZIP: 22202

```

```

; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: Patent Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/235, 277

```

```

; FILING DATE: 29-APR-1994

```

```

; CLASSIFICATION: A36

```

```

; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Oblon, NO. 5733543man F.

```

```

; REGISTRATION NUMBER: 24,618

```

```

; REFERENCE/DOCKET NUMBER: 6042-008-68

```

```

; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (703) 413-3000

```

```

; TELEFAX: (703) 413-2220

```

```

; TELEEX: 248855 OPAT UR

```

```

; INFORMATION FOR SEQ ID NO: 1:

```

```

; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 5653 base pairs

```

```

; TYPE: nucleic acid

```

```

; STRANDEDNESS: single

```

```

; TOPOLOGY: unknown

```

```

; MOLECULE TYPE: Other nucleic acid

```

```

; US-08-235-277-1

```

```

alignment_block:

```

```

US-09-471-276-831_COPY_1_16 x US-08-235-277-1  ..

```

```

Align seg 1/1  to: US-08-235-277-1  from: 1  to: 5653

```

```

      8 leuLeuLeuTrpGlyValThrTrpGly 16
      :::::::::::|||||  |||||
1286 ATTCATATCTCGGGCGGCTCGCGCGC 1312

```

```

alignment_scores:
  Quality: 42.00      Length: 9
  Ratio: 5.250        Gaps: 0
  Percent Similarity: 88.889  Percent Identity: 66.667

```



```
232 AAGAAACAGCCAGATTCTCCATCCCATGACAGACACCATCGGG 281
      |||  ::|||  ::|  |||::|
89 yAgtYrAgtGysArgSerGlyLeuSerThrlGlyTrp**GlnLeuSerL 106
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|
282 GAGATACCCCTGCCTACTATTACAGCTCTGCAGGCTGTGACAGCCGCG 331
106 yslLeuLcGluLeu 110
      |||::|::|::|
332 ACCCCCTGGAGCTG 345

seq_name: /cgn2_6/plodata/2/lna/6B_COMB.seq:us-09-345-468-11

seq_documentation_block:
; Sequence 11, Application US/09345468
; Patent No. 6245527
; GENERAL INFORMATION:
; APPLICANT: Busfield, S.
; APPLICANT: Villevall, J.
; APPLICANT: Jandrot-Perrus, M.
; APPLICANT: Valchenker, W.
; TITLE OF INVENTION: GLYCOPROTEIN VI AND USES THEREOF
; FILE REFERENCE: 7853-147
; CURRENT APPLICATION NUMBER: US/09/345,468
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 2170
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-345-468-11
```

```
alignment_scores:
      Quality: 87.50      Length: 138
      Ratio: 1.509      Gaps: 6
Percent Similarity: 42.029      Percent Identity: 28.261
```

alignment_block:

US-09-471-276-831 x US-09-345-468-11 ..

Align seg 1/1 to: US-09-345-468-11 from: 1 to: 2170

```
1 MeSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
      |||::|::|::|::|::|::|::|::|::|::|::|::|
50 ATGACGCCGCCCTGCACAGCCCTGCTCTGCTGCGCTGAGCTGCGCCC 99
17 oValThGlnAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 34
      |||::|::|::|::|::|::|::|::|::|::|::|::|
100 CAGACCCCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149
34 LuSerGlnHis**LeuLysThrLeuLcGlyncysAspAlaSerPro 50
      |||
150 AG..... 151
51 GlyProProGlyAspSerArgLeuProAlaValGlnGluTrpGly.... 65
      |||::|::|::|::|::|::|::|::|::|::|::|::|
152 .....CCAGGC.....TCTGTATCATCAGTGGGGGAGGCC 180
66 .....AlaGlnGlu..... 68
181 CGTGACCATCTGCTGTACGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 230
69 .....ProValHisLeuAsp.....SerPro 75
231 ATAAAGAGGAGAGCCAGAGCCCTTGACAGAAATAACCACTGGAACCC 280
76 AlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaI 89
      |||::|::|::|::|::|::|::|::|::|::|::|::|
281 AAGACAGAGGAGCATCTCTCCATCCCATGACAGACACCATCGCGG 330
89 yAgtYrAgtGysArgSerGlyLeuSerThrlGlyTrp**GlnLeuSerL 106
```

```
331 GAGATACCCCTGCCTACTATTACAGCTCTGCAGGCTGTGACAGCCGCG 380
      |||::|::|::|::|::|::|::|::|::|::|::|::|
106 yslLeuLcGluLeu 110
      |||::|::|::|
381 ACCCCCTGGAGCTG 394
```

seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq:us-08-985-950-11

seq_documentation_block:

; Sequence 11, Application US/08985950

; Patent No. 6140076

; GENERAL INFORMATION:

; APPLICANT: Adema, Gosse Jan

; TITLE OF INVENTION: Isolated Mammalian Monocyte Cell Genes;

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; STREET: 901 California Avenue

; CITY: Palo Alto

; STATE: California

; COUNTRY: USA

; ZIP: 94304-1104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/985,950

; FILING DATE: 05-DEC-1997

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 60/041,279

; FILING DATE: 21-MARCH-1997

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 60/033,181

; FILING DATE: 16-DEC-1996

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 60/032,252

; FILING DATE: 06-DEC-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Ching, Edwin P.

; REGISTRATION NUMBER: 34,090

; REFERENCE/DOCKET NUMBER: DX0670K

; TELEPHONE: (650)852-9196

; TELEFAX: (650)496-1204

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1620 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 81..1397

; US-08-985-950-11

```
alignment_scores:
      Quality: 80.50      Length: 121
      Ratio: 1.238      Gaps: 7
Percent Similarity: 53.719      Percent Identity: 32.231
```

alignment_block:

US-09-471-276-831 x US-08-985-950-11 ..

Align seg 1/1 to: US-08-985-950-11 from: 1 to: 1620

```
1 MeSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
      |||::|::|::|::|::|::|::|::|::|::|::|::|
```

```
81 ATGACCCCACTCTCAGCGTCTGATCTGTCTGGCGCTGAGCCTGGAGCC 130
17 ova1Thrg1ua1a1a1lePhetYrGlUthrgIn**SerLeuTrpAlaG 34
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 CAGGACCCAGCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 180
34 lUsErGlUhts***LeuYsThrLeuGly.GlnCysAspAlaAspValP 50
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 AGCAGAGCTCTGTGATC...ACCCAGAGGAGGAGGAGGAGGAGGAGGAG 227
50 roGlyPrProGlyAspSerArgLeuProAlaVal.GlnGlu..... 63
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
228 CAGGGAGGAGCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 277
64 .....TrrpGlyAlaGlnGluProValHsLeuAspSerProAlaIle 78
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
278 AGCAGCTCTGATTTACAGGATCTCCACAGAGGCTT.....GTCAAG 318
78 ysh1sGlnPhe.....LeuLeuThrg1YAspThrg1nGlyArgTyr 91
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
319 AGGCGCAGTTCCTCCATCTATCATCATCAGCTGGGAGACATGAGGGGG 368
92 ArgCysArgSerGlyLeuSerThr..GlyTrp**GlnLeuSerLysLe 107
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
369 TGCTGTATCTATGCGAGCCAGCACTGCGAGGCTCTCAGAGAGAGTGA 418
107 uleuGluLeu 110
419 CCGGAGCTG 428
```

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-135-782-3

seq_documentation_block:

```
; Sequence 3, Application US/09135782
; Patent No. 6027929
; GENERAL INFORMATION:
; APPLICANT: Xu, Shuang-Yong
; TITLE OF INVENTION: Method For Cloning And Producing The Napi Restriction
; TITLE OF INVENTION: Endonuclease In E. coli And Purification Of The
; TITLE OF INVENTION: Recombinant Napi Restriction Endonuclease
; FILE REFERENCE: NEB-143
; CURRENT APPLICATION NUMBER: US/09/135,782
; CURRENT FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 735
; TYPE: DNA
; ORGANISM: No. 6027929loc sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(732)
US-09-135-782-3
```

alignment_scores:

Quality:	75.50	Length:	56
Ratio:	2.097	Gaps:	4
Percent Similarity:	64.286	Percent Identity:	39.286

alignment_block:

US-09-471-276-831 x US-09-135-782-3 ..

Align seg 1/1 to: US-09-135-782-3 from: 1 to: 735

```
55 AspSerArgLeuProAlaValGlnGluTrpGlyAlaGlnGluProValH 71
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
394 GATCATAGATTTCG...ATGGAAGATGGGAGCTAGTGAAGCTCCACA 440
71 uleuAspSer.....ProAlaIleYsh1sGlnPhe..LeuLeuT 84
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
441 CTTCACATCTATGATGATTAATCAAGCAAGCAAGTTTCATTTGTA 490
```

```
84 hrcGlyAspThrg1nGlyArgTyrArg.....CysArg 94
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
491 AGAAGATACCTCTGGCATTCACATCTTTAAATCAGAGAGCTGTGAG 540
95 SerGlyLeuSerThrGly 100
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
541 CGATGTATCAAAACTCGT 558
```

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-172-108-48

seq_documentation_block:

```
; Sequence 48, Application US/09172108
; Patent No. 6160104
; GENERAL INFORMATION:
; APPLICANT: Cunningham, Mary Jane
; APPLICANT: Zweiger, Gary B.
; APPLICANT: Panzer, Scott R.
; APPLICANT: Selhammer, Jeffrey J.
; TITLE OF INVENTION: MARKERS FOR PEROXISOMAL PROLIFERATORS
; FILE REFERENCE: PA-0012 US
; CURRENT APPLICATION NUMBER: US/09/172,108
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 48
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 700607235H1
US-09-172-108-48
```

alignment_scores:

Quality:	74.50	Length:	65
Ratio:	1.863	Gaps:	3
Percent Similarity:	61.538	Percent Identity:	36.923

alignment_block:

US-09-471-276-831 x US-09-172-108-48 ..

Align seg 1/1 to: US-09-172-108-48 from: 1 to: 285

```
1 MetSerMetLeuValAlaPheLeuLeuTrpGlyValThrTrpLysPr 17
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
14 ATGTCTGTGCTGACTGACTGACTGACTGCTGTGAGGTTTCATTCTGGGCC 63
17 ova1Thrg1ua1a1a1lePhetYrGlUthrgIn**SerLeuTrpAlaG 34
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
64 AGCACTGACAGACAGCCTGTATATT.....CAAGGAGCCTCGAAGAAC 106
34 lUsErGlUhts***LeuYsThrLeuGlyGlnCysAspAla..... 47
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107 AGTCG.....CTTGCCAGAGGCGCTTGCTTCCCAATCCAGTCCCTG 150
48 .....AspValProGlyProGlyAspSer 56
1 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 GATCAGACCTGGCGCTGAGGAGCATTTCTGCTGTGCGCAGGGAGACAGT 195
```

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-172-711-45

seq_documentation_block:

```
; Sequence 45, Application US/09172711
; Patent No. 6160105
; GENERAL INFORMATION:
; APPLICANT: Cunningham, Mary Jane
; APPLICANT: Zweiger, Gary B.
; APPLICANT: Panzer, Scott R.
; APPLICANT: Selhammer, Jeffrey J.
; TITLE OF INVENTION: MONITORING TOXICOLOGICAL RESPONSES
; FILE REFERENCE: PA-0011 US
; CURRENT APPLICATION NUMBER: US/09/172,711
; CURRENT FILING DATE: 1998-10-13
```

```
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PERL Program
; SEQ ID NO 45
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 700607235H1
US-09-172-711-45
```

```
alignment_scores:
  Quality: 74.50      Length: 65
  Ratio: 1.863      Gaps: 3
  Percent Similarity: 61.538      Percent Identity: 36.923
```

alignment_block:
US-09-471-276-831 x US-09-172-711-45 ..

Align seg 1/1 to: US-09-172-711-45 from: 1 to: 285

```
1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
14 ATCTCTCTGCTGACTGACTGACTGCTGCTGCTGCTGCTGCTGCTGCTG 63
17 cValThrGluAlaAlaIlePheTyrgluThrGln**SerLeuTrpAlaG 34
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 AGCAAGTGCACAGCGCTGTATATT.....CAAGAAAGCCTCGGAAAAC 106
34 InuSerGluHis**LeuLysThrLeuGlyGlnCysAspAla..... 47
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 AGTCC.....CTGCCAGCGCCCTGGCTTCTCTGCAATCCAGTGCCTG 150
48 .....AspValProGlyProProGlyLysSer 56
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 GATCAGACCTGCGCTGAGCATTCCTGCTGCTGCGCAGCGGACACT 195
```

seq_name: /cgn2_6/plodata/2/lna/5B_COMB.seq:US-08-641-038A-1

```
seq_documentation_block:
; Sequence 1, Application US/08641038A
; Patent No. 5861154
; GENERAL INFORMATION:
; APPLICANT: SODA, KENJI
; APPLICANT: TANAKA, HIDEHIKO
; APPLICANT: INOUE, HIROYUKI
; APPLICANT: INAGAKI, KENJI
; APPLICANT: ESAKI, NOBUYOSHI
; TITLE OF INVENTION: RECOMBINANT L-METHIONINE GAMMA-LYASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/641,038A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0020-3989
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
```

```
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 61..1254
US-08-641-038A-1
```

```
alignment_scores:
  Quality: 74.50      Length: 117
  Ratio: 1.242      Gaps: 6
  Percent Similarity: 51.282      Percent Identity: 29.060
```

alignment_block:
US-09-471-276-831 x US-08-641-038A-1 ..

Align seg 1/1 to: US-08-641-038A-1 from: 1 to: 1320

```
2 SerMetLeuValAlaPheLeuLeuLeuTrpGlyValThr.....Trp.G 16
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:
704 GCCATGGCGCATACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 753
16 lYProValThrGluAlaAlaIlePheTyrgluThrGln**SerLeuTrp 32
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:
754 GACCGTATAGCTGTGCGAGGCGCTCAAGACATGACCCGTCGCTGCTTC 803
33 AlGluSerGluHis**LeuLysThrLeuGlyGlnCysAspAlaAspVa 49
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:
804 GCCCATGACGCGCGACTGTGATGCGCGCATCAAGACCTCAACCTGCG 853
49 lProGlyProPro.....GlyAspSerArgLeuProAlaVal 61
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:
854 GCATGACCGCGCGACTGCGCGCAACGCTGCGTGTGCGGAGTTCTGCGC 903
61 aGlnGluTrpGlyAlaGlnGluPro..... 69
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:
904 CGGCAGCGCGCGAGTGTGATTCATACCGCGGCTGCGGACCTTCCC 953
70 ..ValHis.....LeuAspSerProAlaIleLysHisGlnPh 81
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:
954 GCAGTACACCTGCGCGCGCAGCATGAGCAGCGCGGCGGCGATGATCG 1003
81 eLeuLeuThrGlyAspThrGlnGlyArgTyrgCysArgSerGlyLeu 97
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1004 CCTT.....CGAACTCAAGCGCGCATGCTGCGCGCGCGCGCTT 1043
```

seq_name: /cgn2_6/plodata/2/lna/5B_COMB.seq:US-09-059-178-1

```
seq_documentation_block:
; Sequence 1, Application US/09059178
; Patent No. 5863788
; GENERAL INFORMATION:
; APPLICANT: SODA, KENJI
; APPLICANT: TANAKA, HIDEHIKO
; APPLICANT: INOUE, HIROYUKI
; APPLICANT: INAGAKI, KENJI
; APPLICANT: ESAKI, NOBUYOSHI
; TITLE OF INVENTION: RECOMBINANT L-METHIONINE GAMMA-LYASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/059,178
APPLICATION NUMBER: US/09/059,178
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/641,038
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0020-3989
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: CDS
LOCATION: 61..1254
US-09-059-178-1

alignment_scores:
Quality: 74.50 Length: 117
Ratio: 1.242 Gaps: 6
Percent Similarity: 51.282 Percent Identity: 29.060

alignment_block:
US-09-471-276-831 x US-09-059-178-1 ..

Align seg 1/1 to: US-09-059-178-1 from: 1 to: 1320

```
2 SerMetLeuValValPheLeuLeuLeuTrpGlyValThr.....Trp.G 16
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
704 GCCATGGCCGACATCACTGCTGTCATGTCGTGGGAGCCGACCTGCTGC 753
16 lProValThrGluAlaAlaIlePheThrGluThrGln**SerLeuTrp 32
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
754 GACCATATACGCTGCGACGGCCCTCAAGACATGACCGGTCGCTGCTGC 803
33 AlaGluSerGluHis**LeuLysThrLeuGlyGlnCysAspAlaAspVa 49
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
804 GCCCATGACGGCCGACACTGTTGATGCGCGGACATCAAGACCTCAACCTGC 853
49 lProGlyProPro.....GlyAspSerArgLeuProAlaVal 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
854 GCATGACACCCGCTGCGCCAGCATGAGCTCAGGTGCTGCGGAGTTCTGCGC 903
61 AlGlnGluTrpGlyAlaGlnGluPro.....GlyAspSerArgLeuProAlaVal 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
904 CGGACCGCGGAGCTGACGCTGATTCATACCGGGCTGCGGACCTTCCC 953
70 ...ValHis.....LeuAspSerProAlaIleLysHisGlnPro 81
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
954 GCATGACACCCGCTGCGCCAGCATGAGCTCAGGTGCTGCGGAGTTCTGCGC 1003
81 GluLeuThrGlyAspThrGlnGlyArgGlyArgSerGlyLeu 97
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1004 CATT.....CGAAGCTCAAGCGCGCATGCGTCCGCGGCGGCTT 1043
```

seq_name: /gcn2_6/ptodata/2/lna/5B_COMB.seq:US-08-642-541-1

seq_documentation_block:
; Sequence 1, Application US/08642541
; Patent No. 5891704

GENERAL INFORMATION:
APPLICANT: YUYING, TAN
TITLE OF INVENTION: HIGH EXPRESSION MODULES ENCODING
TITLE OF INVENTION: METHIONINASE, PURIFIED RECOMBINANT METHIONINASE PRODUCED
TITLE OF INVENTION: METHIONINASE, PURIFIED RECOMBINANT METHIONINASE PRODUCED
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,541
FILING DATE: 03-MAY-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: MILLMAN, Robert A.
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 31276-20002.21
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1369 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 48..1241
US-08-642-541-1

alignment_scores:
Quality: 74.50 Length: 117
Ratio: 1.242 Gaps: 6
Percent Similarity: 51.282 Percent Identity: 29.060

alignment_block:
US-09-471-276-831 x US-08-642-541-1 ..

Align seg 1/1 to: US-08-642-541-1 from: 1 to: 1369

```
2 SerMetLeuValValPheLeuLeuLeuTrpGlyValThr.....Trp.G 16
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
691 GCCATGGCCGACATCACTGCTGTCATGTCGTGGGAGCCGACCTGCTGC 740
16 lProValThrGluAlaAlaIlePheThrGluThrGln**SerLeuTrp 32
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
741 GACCATATACGCTGCGACGGCCCTCAAGACATGACCGGTCGCTGCTGC 790
33 AlaGluSerGluHis**LeuLysThrLeuGlyGlnCysAspAlaAspVa 49
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
791 GCCCATGACCGCCGACACTGTTGATGCGCGGACATCAAGACCTCAACCTGC 840
49 lProGlyProPro.....GlyAspSerArgLeuProAlaVal 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
841 GCATGACACCGCTGCGCCAGCATGAGCTCAGGTGCTGCGGAGTTCTGCGC 890
61 AlGlnGluTrpGlyAlaGlnGluPro.....GlyAspSerArgLeuProAlaVal 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
891 CGGACCGCGGAGCTGACGCTGATTCATACCGGGCTGCGGAGCTTCCC 940
```


alignment_scores:
 Quality: 74.50 Length: 117
 Ratio: 1.242 Gaps: 6
 Percent Similarity: 51.282 Percent Identity: 29.060

alignment_block:

US-09-471-276-831 x US-09-260-889-1 ..

Align seg 1/1 to: US-09-260-889-1 from: 1 to: 1369

```

2 SerMetLeuValValPheLeuLeuLeuTrrpGlyValThr.....TrrpG 16
   :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
691 GCGATGCGCCAGCATCTGCTGCTGCTGCTGCGCAGCCAGCCAGCATCTG 740
   ||||| ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
16 lYpValThrGluAlaAlaIlePheTyrGluThrGln**SerLeuTrrp 32
   ||||| ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
741 GACCGTATACGTCTGCAGGGCCCTCAAGACATGACCGGTGCGTCTC 790
   ||||| ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
33 AlaGluSerGluHis**LeuIysThrLeuGlyGlnCysAspAlaAspVa 49
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
791 GCGCCATGACGCCGACACTGTGATGCGCGGCATCAAGACCCCTACACCTG 840
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
49 lProGlyProPro.....GlyAspSerArgLeuProAlaVal 61
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
841 GCATGACCGCCAGCTGCGCCAGCAGCTCAGTCTGCTGCGGATCTCTGCC 890
   ||||| ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
61 AlGlnGluTrrpGlyAlaGlnGluPro.....GlyAspSerArgLeuProAlaVal 69
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
891 CGCGACCCCGCAGGTGGAGCTGATCCATTCACCGCGGCTGCGGAGCTTCCC 940
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
70 ...ValHis.....LeuAspSerProAlaIleIysHisGlnPh 81
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
941 GCAGTACACCCCTGCGCCGCGCAGCAGTGAAGCCAGCCGCGCGCATGTCG 990
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
81 eLeuLeuThrGlyAspThrGlnGlyArgTyrArgCysArgSerGlyLeu 97
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
991 CCTT.....CGAAGTCMAAGCGCGCATCGTCTGCGCGCGCGCTT 1030
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-260-889-3

seq_documentation_block:

```

: Sequence 3, Application US/09260889
: Patent No. 6231854
: GENERAL INFORMATION:
: APPLICANT: TAN, Yuying
: TITLE OF INVENTION: HIGH EXPRESSION MODULES ENCODING
: TITLE OF INVENTION: METHIONINASE
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 2000 Pennsylvania Avenue, NW, suite 5500
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20006-1868
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows
: SOFTWARE: FASTSEQ for Windows Version 2.0b
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/260.889
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/914,377
: FILING DATE: 19-AUG-1997
: APPLICATION NUMBER: 08/642,541
: FILING DATE: 03-MAY-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Murashige, Kale H
: REGISTRATION NUMBER: 29,959
: REFERENCE/DOCKET NUMBER: 31276-20002.11

```

TELECOMMUNICATION INFORMATION:

```

: TELEPHONE: 202-887-1500
: TELEFAX: 202-822-0168
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1369 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: ANTI-SENSE: YES
: US-09-260-889-3

```

alignment_scores:
 Quality: 74.50 Length: 117
 Ratio: 1.242 Gaps: 6
 Percent Similarity: 51.282 Percent Identity: 29.060

alignment_block:

US-09-471-276-831 x US-09-260-889-3/rev ..

Align seg 1/1 to reverse of: US-09-260-889-3 from: 1 to: 1369

```

2 SerMetLeuValValPheLeuLeuLeuTrrpGlyValThr.....TrrpG 16
   :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
679 GCGATGCGCCAGCATCTGCTGCTGCTGCTGCGCAGCCAGCCAGCATCTG 630
   ||||| ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
16 lYpValThrGluAlaAlaIlePheTyrGluThrGln**SerLeuTrrp 32
   ||||| ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
629 GACCGTATACGTCTGCAGGGCCCTCAAGACATGACCGGTGCGTCTCTC 580
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
33 AlaGluSerGluHis**LeuIysThrLeuGlyGlnCysAspAlaAspVa 49
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
579 GCGCCATGACGCCGACACTGTGATGCGCGGCATCAAGACCCCTACACCTG 530
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
49 lProGlyProPro.....GlyAspSerArgLeuProAlaVal 61
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
529 GCATGACCGCCAGCTGCGCCAGCAGCTCAGTCTGCTGCGGATCTCTGCC 480
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
61 AlGlnGluTrrpGlyAlaGlnGluPro.....GlyAspSerArgLeuProAlaVal 69
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
479 CGCGACCCCGCAGGTGGAGCTGATCCATTCACCGCGGCTGCGGAGCTTCCC 430
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
70 ...ValHis.....LeuAspSerProAlaIleIysHisGlnPh 81
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
429 GCAGTACACCCCTGCGCCGCGCAGCAGATGAAGCCAGCCGCGCGCATGTCG 380
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
81 eLeuLeuThrGlyAspThrGlnGlyArgTyrArgCysArgSerGlyLeu 97
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
379 CCTT.....CGAAGTCMAAGCGCGCATCGTCTGCGCGCGCGCTT 340
   ||||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-384-106A-8

seq_documentation_block:

```

: Sequence 8, Application US/08384106A
: Patent No. 6033847
: GENERAL INFORMATION:
: APPLICANT: Sherr Ph.D., Charles J.
: APPLICANT: Downing M.D., James
: APPLICANT: Hlrel Ph.D., Hiroshi
: APPLICANT: Okuda, Tsukasa
: TITLE OF INVENTION: Ink4c-p18 and Ink4d-p19, Inhibitors of
: TITLE OF INVENTION: Cyclin Dependent Kinases Cdk4 and Cdk6, and Uses Thereof
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSSTEIN & FOX
: STREET: 1100 New York Ave., N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005

```



```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,106A
FILING DATE: 06-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fox, Samuel L.
REGISTRATION NUMBER: 30,353
REFERENCE/DOCKET NUMBER: 0656,0500000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ. ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 501 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..501
US-08-384-106A-8

```

```

alignment_scores:
Quality: 74.00      Length: 99
Ratio: 1.542        Gaps: 6
Percent Similarity: 48.485   Percent Identity: 30.303

```

alignment_block:
US-09-471-276-831 x US-08-384-106A-8 ..

Align seg 1/1 to: US-08-384-106A-8 from: 1 to: 501

```

17 ProValThrGluAlaAlaIle.....PhenylGluThrGln***SerLe 31
||||| ::||| ||| ::|||
229 CCAATTCATGACGACGCCGACCTGCTCTGACAC..... 267
31 uTPAlaGluSerGluHis**LeuLysThrLeuGlyGlnCysAspAla 48
||||| ::||| ||| ::|||
268 .....CTGAAGTCTCTAGTGAGCAGCGGGCTG 295
48 sPValProGlyProProGlyAspSerArgLeuPro.....AlaVal 61
||||| ||| ||| ::||| |||||
296 ATGTCACAGTGCCTGATGCGACCGGGCAGCTTCCAATCCATCTGCGACTT 345
62 GlnGlu.....TrpGlyAlaGlnGluProVa 70
||||| ::||| ||| ::|||
346 CAGAGAGGCTCACACTGCTGTGTCAGCTTCTGTGACAGCTGAATCTGATCT 395
70 LHisLeuAspSerProAlaIleLysHisGlnPheLeuLeuThrGlyAsp 86
||||| ::||| ||| ::|||
396 CCATTCGACGAGGACGCGGCTCTCACACCTTGG..... 430
87 ThrGlnGlyArgTyrArgCysArgSerGlyLeuSerThrGlyTrp 101
||||| ::||| ||| ::|||
431 .....AGCTGCACTGACGAGGCGCTCACACCTTGG 466

```

seq_name: /cgn2_6/ptodata/2/lna/PCTUS_COMB.seq: PCT-US96-01643-8

```

seq_documentation_block:
; Sequence 8, Application PC/TUS9601643
; GENERAL INFORMATION:
; APPLICANT: St. Jude Children's Research Hospital
; TITLE OF INVENTION: InK4c-p18 and InK4d-p19, Inhibitors of
; NUMBER OF SEQUENCES: 15
; CYCLIN Dependent Kinases CDK4 and CDK6, and uses thereof
; CORRESPONDENCE ADDRESS:

```

```

ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 New York Ave., N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/01643
FILING DATE: 06-FEB-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/384,106
FILING DATE: 06-FEB-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fox, Samuel L.
REGISTRATION NUMBER: 30,353
REFERENCE/DOCKET NUMBER: 0656,050PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ. ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 501 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
POSITION IN GENOME:
CHROMOSOME/SEGMENT: p19
FEATURE:
NAME/KEY: CDS
LOCATION: 1..501
PCT-US96-01643-8

```

```

alignment_scores:
Quality: 74.00      Length: 99
Ratio: 1.542        Gaps: 6
Percent Similarity: 48.485   Percent Identity: 30.303

```

alignment_block:
US-09-471-276-831 x PCT-US96-01643-8 ..

Align seg 1/1 to: PCT-US96-01643-8 from: 1 to: 501

```

17 ProValThrGluAlaAlaIle.....PhenylGluThrGln***SerLe 31
||||| ::||| ||| ::|||
229 CCAATTCATGACGACGCCGACCTGATCTCTGACAC..... 267
31 uTPAlaGluSerGluHis**LeuLysThrLeuGlyGlnCysAspAla 48
||||| ::||| ||| ::|||
268 .....CTGAAGTCTCTAGTGAGCAGCGGGCTG 295
48 sPValProGlyProProGlyAspSerArgLeuPro.....AlaVal 61
||||| ||| ||| ::||| |||||
296 ATGTCACAGTGCCTGATGCGACCGGGCAGCTTCCAATCCATCTGCGACTT 345
62 GlnGlu.....TrpGlyAlaGlnGluProVa 70
||||| ::||| ||| ::|||
346 CAGAGAGGCTCACACTGCTGTGTCAGCTTCTGTGACAGCTGAATCTGATCT 395
70 LHisLeuAspSerProAlaIleLysHisGlnPheLeuLeuThrGlyAsp 86
||||| ::||| ||| ::|||
396 CCATTCGACGAGGACGCGGCTCTCACACCTTGG..... 430

```



```

450 CCATGCAGAGCAGCCAGGGGTCTCACACCTTGG..... 484
87 ThuringiarytTargCySarGserglyLeuSerThgIyTrp 101
      :::::|||||:::|||||  |||:::|
485 .....AGCTGGCACYGCAGAGGGGCTCAGAGACTCGTGG 520

seq_name: /cgn2_6/ptodata/2/ina/PCtus_comb.seq:PCt-US96-01643-7

seq_documentation_block:
; Sequence 7, Application PC/TUS9601643
; GENERAL INFORMATION:
; APPLICANT: St. Jude Children's Research Hospital
; TITLE OF INVENTION: InK4C-p18 and InK4d-p19, Inhibitors of
; TITLE OF INVENTION: Cyclin dependent kinases CDK4 and CDK6, and Uses Thereof
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
;

```

seq.documentation.block:
: Sequence 7, Application PC/TUS9601643
: GENERAL INFORMATION:
: APPLICANT: St. Jude Children's Research Hospital
: TITLE OF INVENTION: Ink4c-p18 and Ink4d-p19, Inhibitors of
: NUMBER OF INVENTION: Cyclin Dependent Kinases CDK4 and CDK6, and Uses Thereof
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
: STREET: 1100 New York Ave., N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/01643
: FILING DATE: 06-FEB-1996
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/384,106
: FILING DATE: 06-FEB-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Fox, Samuel L.
: REGISTRATION NUMBER: 30,353
: REFERENCE/DOCKET NUMBER: 0656, 050PC01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-2600
: TELEFAX: 202-371-2540
: INFORMATION FOR SEQ ID NO: 7:
:

```

?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 555 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: cdna
?

```

```

; ORGANISM: Homo sapiens
;
; POSITION IN GENOME:

```

CHROMOSOME/SEGMENT: p19
PCT-US96-01643-7

```

alignment_scores:
    quality: 74.00      length: 99
    ratio: 1.542      gaps: 6
    percent similarity: 48.485      percent identity: 30.303

alignment_block:
US-09-471-276-831 x PCR-US96-01643-7 ..

Align seg 1/1 to: PCT-US96-01643-7 from: 1 to: 555

17 ProvalThrgJualAlaIle.....PheYrGluThrgIn**SerLe 31
||||| ::||| ::|||
283 CCAgTCcATGAAGcAGcCCGcCATGgATTCCTGcAGAcC..... 322
31 uTrrpLacluserGIuHs***LeuYsThrLeuGlycIncYaspaIaa 48
||||| ::||| ::|||
322 .....CTCAAGCTCTCGTGGTAGGcAGCGGGcCTG 349

```

```

48  spval1ProGly1ProProGly1aspSerArgLeuPro.....AlaVal 61
      |||||  |||||  |||||  |||||  |||||  |||||
350  ATCTCAACGCTGCTATGTGGCACCGGGGCACTTCCAAATCATCTGGCAGTT 3939
      |||||  |||||  |||||  |||||  |||||  |||||
62  GInGlu.....TPrp1ValagInGluProVa 70
      |||||  |||||  |||||  |||||  |||||  |||||
400  CAAGACGGCACACTGCTGTGTCAGCTTTCGCGACAGTGAATCTGATCT 4449
      |||||  |||||  |||||  |||||  |||||  |||||
70  IHis..LeuaspSerProAlaIleIleYshIagInPheLeuLeuThrGlyasp 86
      |||||  |||||  |||||  |||||  |||||  |||||
450  CCATCGCAGCGCAGCGCACGGGCTTCACACCCCTTGG.....484
      |||||  |||||  |||||  |||||  |||||  |||||
87  ThrInGInIArgTArgCysArgSerGlyLeuSerThrGlyTrp 101
      |||||  |||||  |||||  |||||  |||||  |||||
485  .....AGCTGGCACTGCAGAGAGGGGCTTCAGACACTGTGTGG 520

```

```
seq_documentation_block:
: Sequence 1, Application PC/TUS9605252
```

TITLE OF INVENTION: p19: A Cell Cycle Inhibitor
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew
 STREET: One Market Plaza, Stewart Street Tower
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94105-1492
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.300
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/05252
 FILING DATE: Not yet assigned
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/425,093
 FILING DATE: 17-APR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Weber, Ellen L.
 REGISTRATION NUMBER: 32,762
 REFERENCE/DOCKET NUMBER: 02307B-059910PC
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 543-9600
 TELEFAX: (415) 543-5043
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 706 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..498
 OTHER INFORMATION: /product= "human p19"
 PCT-US96-05252-1

```
alignment_scores:
  Quality: 74.00
  Ratio: 1.542
  Percent Similarity: 48.485
  Length: 99
  Gaps: 6
  Percent Identity: 30.303
```

alignment_block;
US-09-471-276-B31 x PCT-US96-05252-1

```

Align seg 1/1 to: PCT-US96-05252-1 from: 1 to: 706

17 ProValThrGluAlaIle.....PheTyrGluThrGln***SerLe 31
||||| :|||
||||| :|||
229 CCAGTCGACGTGACGACGCCGACGTGATTCCTGGACACC..... 267
||||| :|||

31 uTrpIaGluSerGluHis***LeuLysThrLeuGlyGlnCysAspAla 48
||||| :|||
||||| :|||

268 .....CTGAAGGTCCTTAGTGGACACCGGGCTG 295

48 sPvalProGlyProProGlyAspSerAlaLeuPro.....AlaVal 61
||||| ||| ||| :|||
||||| :|||

296 AATGTCACGCTGCGTATGATGCACCGGGGACGCTCCAAATGCAGCGATT 345
||||| :|||

62 GInGlu.....TrpGlyAlaGlnIupProva 70
||||| :|||
||||| :|||

346 CAAAGGAGGTCACACTGCTGTGTGTCAGCTTCTTGCGACACTGAATCTGATCT 395
||||| :|||

70 LHisLeuAspSerProAlaIleLysHisGlnPheLeuLeuThrGlyAsp 86
||||| :|||
||||| :|||

396 CCATGCGCAGGAGCGCCGACGGGTCTTCACACCCCTTGG..... 430
||||| :|||

87 ThrGlnGlyArgTyrArgCysArgSerGlyLeuSerThrGlyTrp 101
||||| :|||
||||| :|||

431 .....ACGTGGCACTGCGACAGAGGGGCTCAGACACTCTGTGG 466
||||| :|||
||||| :|||

seq_name: /cgn2_6/prodata/2/1aa/5b_comb_seg: US-08-667-939A-13

```

seq_documentation_block:

```

1 PatentNo. 5998166
2 General INFORMATION:
3 APPLICANT: LDO, Shun
4 TITLE OF INVENTION: CD16-II VARIANTS
5 NUMBER OF SEQUENCES: 25
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: BROWDY AND NEIMARK
8 STREET: 419 Seventh Street, N.W., Suite 300
9 CITY: Washington
10 STATE: D.C.
11 COUNTRY: USA
12 ZIP: 20004
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: PatentIn Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/667,939A
20 FILING DATE: 24-JUN-1996
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US 08/433,123
23 FILING DATE: 03-MAY-1995
24 ATTORNEY/AGENT INFORMATION:
25 NAME: BROWDY, Roger L.
26 REGISTRATION NUMBER: 25,618
27 REFERENCE/DOCKET NUMBER: LDO-2A
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: 202-628-5197
30 TELEFAX: 202-737-3528
31 INFORMATION FOR SEO ID NO: 13:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 765 base pairs
34 TYPE: nucleic acid
35 STRANDEDNESS: single
36 TOPOLOGY: linear
37 MOLECULE TYPE: cDNA
38
39 US-08-667-939A-13

```

alignment_scores:	
Quality:	74.00
Ratio:	1.423
Length:	91
Gaps:	7

Percent Similarity: 57.143 Percent Identity: 35.165

alignment_block:

US-09-471-276-831 x US-08-667-939A-13 ..

Align seg 1/1 to: US-08-667-939A-13 from: 1 to: 765

```
43 GYGLNCYSASPAALASPVALPROGLY.....ProProGly.Asps 56
|||||.....|
120 GGACAGTGTACTCTGAAGTCCAGGAGCCTACTCCCTGAGACATTT 169
56 eArGLeuProAlaValAlGInLUtRPLyAlaGInGLuProValHISLeu 72
||.....|
170 CC.....ACACAGTGTTCACAAAGAG.....AACCTC 198
73 AsperProAlaIleLysHISGLNpheLeu.....LeuThrGLYAspTh 87
|||.....|
199 ATCTCAAGCCAGCGCCTGAGCTTCTATGTAGCGCTGCCACAGTCGACGA 248
87 rGLInGLYArGTyRArgCysArgSerGLYLeuSerThrGLYTrp**GInL 104
.....|
249 CAGTGAGAGTACAGGTGCCAGACGAACTCTCCACC.....C 286
104 eUsErLysLeuLeuGLuLeu.ThrGLYProLysValLeuAlaCysSerLe 120
|||||.....|
287 TCAGTGACCCGCGTGACAGTCAAGTCCAGTGGCTGCTCTCTCCAG 336
120 uAlaLeuAspGLYAlaSer 126
|||.....|
337 GCCCTCGTGCGGTGTCA 355
```

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-667-939A-14

seq_documentation_block:

; Sequence 14, Application US/08667939A

; Patent No. 5998166

; GENERAL INFORMATION:

; APPLICANT: LUD, Shun

; TITLE OF INVENTION: CD16-11 VARIANTS

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 Seventh Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; FILING DATE: 24-JUN-1996

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/433,123

; FILING DATE: 03-MAY-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: BROWDY, Roger L.

; REGISTRATION NUMBER: 25,618

; TELEPHONE: 202-628-5197

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-628-5197

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 765 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single.

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; US-08-667-939A-14

alignment_scores:

Quality: 74.00 Length: 91

Ratio: 1.423 Gaps: 7

Percent Similarity: 57.143 Percent Identity: 35.165

alignment_block:

US-09-471-276-831 x US-08-667-939A-14 ..

Align seg 1/1 to: US-08-667-939A-14 from: 1 to: 765

```
43 GYGLNCYSASPAALASPVALPROGLY.....ProProGly.Asps 56
|||||.....|
120 GGACAGTGTACTCTGAAGTCCAGGAGCCTACTCCCTGAGACATTT 169
56 eArGLeuProAlaValAlGInLUtRPLyAlaGInGLuProValHISLeu 72
||.....|
170 CC.....ACACAGTGTTCACAAAGAG.....AACCTC 198
73 AsperProAlaIleLysHISGLNpheLeu.....LeuThrGLYAspTh 87
|||.....|
199 ATCTCAAGCCAGCGCCTGAGCTTCTATGTAGCGCTGCCACAGTCGACGA 248
87 rGLInGLYArGTyRArgCysArgSerGLYLeuSerThrGLYTrp**GInL 104
.....|
249 CAGTGAGAGTACAGGTGCCAGACAACTCTCCACC.....C 286
104 eUsErLysLeuLeuGLuLeu.ThrGLYProLysValLeuAlaCysSerLe 120
|||||.....|
287 TCAGTGACCCGCGTGACAGTCAAGTCCAGTGGCTGCTCTCTCCAG 336
120 uAlaLeuAspGLYAlaSer 126
|||.....|
337 GCCCTCGTGCGGTGTCA 355
```

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-483-376-1

seq_documentation_block:

; Sequence 1, Application US/08483376

; Patent No. 5955330

; GENERAL INFORMATION:

; APPLICANT: Vasil, Vinita

; APPLICANT: Clancy, Maureen A.

; APPLICANT: Ferl, Robert J.

; APPLICANT: Vasil, Indra K.

; APPLICANT: Hannah, L. C.

; TITLE OF INVENTION: No. 5955330el Means for Enhancing Gene

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

; STREET: 5370 Manhattan Circle, Suite 201

; CITY: Boulder

; STATE: Colorado

; COUNTRY: US

; ZIP: 80303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/418,540

; FILING DATE: 07-APR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/102,115

; FILING DATE: 04-AUG-1993

; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/830,956
FILING DATE: 05-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/353,854
FILING DATE: 18-MAY-1989
ATTORNEY/AGENT INFORMATION:
NAME: Farber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 10-948
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6386 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Zea mays
STRAIN: Black Sweet
FEATURE:
NAME/KEY: exon
LOCATION: 131..182
FEATURE:
NAME/KEY: exon
LOCATION: 1211..1324
FEATURE:
NAME/KEY: exon
LOCATION: 1828..1948
FEATURE:
NAME/KEY: exon
LOCATION: 2041..2187
FEATURE:
NAME/KEY: exon
LOCATION: 2269..2460
FEATURE:
NAME/KEY: exon
LOCATION: 2605..2728
FEATURE:
NAME/KEY: exon
LOCATION: 2822..3038
FEATURE:
NAME/KEY: exon
LOCATION: 3256..3351
FEATURE:
NAME/KEY: exon
LOCATION: 3447..3620
FEATURE:
NAME/KEY: exon
LOCATION: 3702..3818
FEATURE:
NAME/KEY: exon
LOCATION: 3912..4078
FEATURE:
NAME/KEY: exon
LOCATION: 4158..4381
FEATURE:
NAME/KEY: exon
LOCATION: 4517..4835
FEATURE:
NAME/KEY: exon
LOCATION: 4768..5212
FEATURE:
NAME/KEY: exon
LOCATION: 5372..5510
FEATURE:
NAME/KEY: exon
LOCATION: 5636..5917
US-08-483-376-1

alignment_scores:
Quality: 73.00 Length: 149
Ratio: 1.106 Gaps: 5
Percent Similarity: 44.295 Percent Identity: 20.805
alignment_block:
US-09-471-276-831 x US-08-483-376-1/rev ..
Align seq 1/1 to reverse of: US-08-483-376-1 from: 1 to: 6386
11 TrpGlyValThrTrpGlyProValThrGluAlaIlePheTyrgluTh 27
||| :||||| ||| ||||| :||| :|||
5643 TGC...CTGCCTGATTCCAGAAAGACGCGCTGCTGCTGACT 5597
27 rgln**SerLeuTrpAlaIleuSerGluHis**Ileu..... 39
:: ||| :||||| :|||
5596 CGTTTAGACTGCTGGAATGGAGACGACGTCGTCCTGACGAGC 5547
39 39
5546 AGAGCAGAGCAGCAGCAGCATCAACGCTTACCGAGCTGACGTA 5497
40LysThrLeuGlyGlnCysAspAlaAspValProGlyPro 54
:: :||| :||| :||| :|||
5496 TCAGGCGCTAGACACATCTCCATTGACGCGCGCTGCGCGCTTCC 5447
54 YAspSerArg,LeuPro...AlaValGlnGluTrpGlyAlaGln 69
| ||||| ||||| ||||| :||| :|||
5446 TTGCTCAGCTACTCTCCAGAACCCCTACACGCGGTGACGATC 5397
70 ValHisLeuAspSerProAlaIleLeuHisGln..... 80
:: :||| :||| :||| :|||
5396 CTCGAGTAGAGCTTCCAGGTGACGTGAGCAATCAACGCTTGA 5347
81PheLeuLeuThrGlyAspThrGlnGlyArgTyr 92
::||| ||||| :||| :|||
5346 CAGTATTTTACTACTCTCTCTACTAGAGAGGTACGCGCATTC 5297
92 rGcYAspArgSerGlyLeuSerThrGlyTrp**GlnLeuSerLys 108
||| :||| :||| :||| :|||
5296 CGAGTGTCCGGGAGTAAACAGATCGCGCTTCAGTCAACAGA 5247
109 GluLeuThrGlyProLysValLeuAlaCysSerLeuAlaLeu 123
||| ||| :||| :||| :|||
5246 ACATTGATGCGCAGAGAGAAAAAATGCTACTTCTCATTAAT 5202
seq_name: /cgn2_6/plodata/2/lna/6A.COMB.seq:US-09-193-191-3
seq_documentation_block:
Sequence 3, Application US/09193191
Patent No. 6130078
GENERAL INFORMATION:
APPLICANT: XIAO, JIAN-PING
TITLE OF INVENTION: METHOD FOR CLONING THE NSPHI RESTRICTION-MODIFICATION
TITLE OF INVENTION: SYSTEM IN E. COLI AND PRODUCING THE RECOMBINANT NSPHI
FILE REFERENCE: NSPHI
CURRENT APPLICATION NUMBER: US/09/193,191
CURRENT FILING DATE: 1998-11-17
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 735
TYPE: DNA
ORGANISM: NO. 6130078loc sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(735)
US-09-193-191-3

Alignment_scores:

Quality: 72.50 Length: 56
 Ratio: 1.959 Gaps: 4
 Percent Similarity: 66.071 Percent Identity: 37.500

Alignment_block:

US-09-471-276-831 x US-09-193-191-3 ..

Align seg 1/1 to: US-09-193-191-3 from: 1 to: 735

```

55 AspSerArgLeuProAlaValGlnGluTrpGlyAlaGlnGluProValH1 71
   ||| |||::||| :::::|||||::|::|::|::|::|::|::|::|::|
394 GATCATCGATCCCG...ATGAAACGTTGGGAGAGCTAGTGCCTCCACA 440
   ||| ||| ::| ::|::|::|::|::|::|::|::|::|::|::|::|
71 sLeuAspSerProAla.....IleLysHisGlnPhe..LeuLeuT 84
   ||| ||| ::| ::|::|::|::|::|::|::|::|::|::|::|::|
441 CTTAACCTTCATGATGATGATGAATTAAGCAAAAGTTTCATTTGTTAA 490
   |||:::|||||::|::|::|::|::|::|::|::|::|::|::|::|
84 hrcGlyAspThrGlnGlyArgTyrArg.....CysArg 94
   |||:::|||||::|::|::|::|::|::|::|::|::|::|::|::|
491 AAAAAGACGATCAGTATACAAATCTTTAAATCGAGAGAGTTGTGAG 540
   |||:::|||||::|::|::|::|::|::|::|::|::|::|::|::|
95 serGlyLeuSerThrGly 100
   :::::|||||
541 CGCTGATCAAAACTGCT 558

```

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-07-668-648-3

seq_documentation_block:

; Sequence 3, Application US/07668648

; Patent No. 5416192

GENERAL INFORMATION:

APPLICANT: Shoyab, Mohammed
 APPLICANT: Plozman, Gregory D.
 TITLE OF INVENTION: EPITHELINS: NOVEL CYSTEINE-RICH GROWTH
 TITLE OF INVENTION: MODULATING PROTEINS
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/668,648
 FILING DATE: 19910819
 CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
 NAME: Mistock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 5624-161-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)790-9090
 TELEFAX: (212) 869-9741
 INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1779 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 TISSUE TYPE: Kidney
 FEATURE:
 NAME/KEY: CDS

; LOCATION: 1..1779
 US-07-668-648-3

alignment_scores:

Quality: 71.50 Length: 140
 Ratio: 1.172 Gaps: 6
 Percent Similarity: 43.571 Percent Identity: 23.571

alignment_block:

US-09-471-276-831 x US-07-668-648-3 ..

Align seg 1/1 to: US-07-668-648-3 from: 1 to: 1779

```

10 LeuTrpGlyValThrTrp.....GlyProValThrGlnAlaVal 23
   ||||| :::::||||| |||::|::| |||::|::|
80 CTGGGGCTCTGCTGACCCCGGAGAGCCAGCTGAGCTGCTGCGGT 129
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|
23 ePheTyrGluThrGln.....**SerLeuTrpAlaGluS 35
   : ::|::|::|::|::|::|::|::|::|::|::|::|::|
130 CCCCTTCTGGACAAATGCCCCACACACTGACGAGCATCTGGTGGCC 179
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
35 erGlnHis**LeuLysThr..LeuGlyGlnCysAspAlaAspValpro 50
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
180 CTGCCAGGTGATGCCCATCTGCTGCCGCCACTCTGCATCTTTACCG 229
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
51 GlyProGlyAspSerArgLeuProAlaValGlnGluTrp..... 64
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
230 TCTCAGGACTTCAGTTGCTGCCCTTCCAGAGCCGCTGCATGGCGG 279
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
65 .....GlyAlaGlnGluProValHisLeuAspS 74
   |||::|::|::|::|::|::|::|::|::|::|::|::|
280 GATGGCCATCCTCTGCCACGGGCTTCACCTGCAGTGCAGAGCGCG 329
   |||::|::|::|::|::|::|::|::|::|::|::|::|
74 erProAlaIleLysHisGlnPheLeuLeuThrGlyAspThrGlnGlyArg 90
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
330 ATCTGCTTCCAAAGATCAG..... 349
   |||::|::|::|::|::|::|::|::|::|::|::|::|
91 TyrArgCysArgSerGlyLeuSerThrGlyTrp**GlnLeuSerLysLe 107
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
350 .....GTAAACACTCTCGGGGTGCCATCCAGTGCCT 381
   |||::|::|::|::|::|::|::|::|::|::|::|::|
107 uLeu.....GluLeuThrGlyProLysValLeuAlaCysSerL 120
   |||::|::|::|::|::|::|::|::|::|::|::|::|
382 GATAGTCACTGCGAATGCCCGGACTTCTCCACGTGCTGTATTATGTGA 431
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
432 TGGCTCCTCGGGGCTGCTGCC 451

```

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-429-998-3

seq_documentation_block:

; Sequence 3, Application US/08429998

; Patent No. 5885961

GENERAL INFORMATION:

APPLICANT: Shoyab, Mohammed
 APPLICANT: Plozman, Gregory D.
 TITLE OF INVENTION: EPITHELINS: NOVEL CYSTEINE-RICH GROWTH
 TITLE OF INVENTION: MODULATING PROTEINS
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/429,998
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/668,648
FILING DATE: 13-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-161-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)790-9090
TELEFAX: (212) 869-9741
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1779 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: kidney
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1779
US-08-429-998-3

```

```

alignment_scores:
Quality: 71.50 Length: 140
Ratio: 1.172 Gaps: 6
Percent Similarity: 43.571 Percent Identity: 23.571

```

alignment_block:

US-09-471-276-831 x US-08-429-998-3 ..

Align seg 1/1 to: US-08-429-998-3 from: 1 to: 1779

```

10 LeuTrpGlyValThrTrp.....GlyProValThrGluAlaAla 23
||||| :||||| |||||
80 CTGTGGCTCTGCTGCTGACCGCGAGGACGACGTAACCTCTCCCT 129
23 ephetyrGluThrGln.....**SerLeuTrpAlaGlu 35
: : : ||||| |||||
130 CCCCTCTGTGACAAATGCGCCACACACAGCAGGATCTGGTGCC 179
35 eGluHis**LeuLysThr...LeuGlyGlnCysAspAlaAspValPro 50
: : : ||||| |||||
180 CTGCGAGGTGATGCCACTGCTGCGGCGCACCTCGCATCTTACCG 229
51 GlyProProGlyAspSerArgLeuProAlaValGlnGluTrp..... 64
: : : ||||| :|||
230 TCTCAGGAGCTTCACGTGCTGCCCTTCCAGAGGCGCGCATGCC 279
65 .....GlyAlaGlnGluProValHisLeuAsp 74
||||| :|||||
280 GATGGGCATCAGCTGCTGCCAGGGGCTTCAGTGCATGCAACAGG 329
74 ePrProAlaIleLysHisGlnPheLeuLeuThrGlyAspThrGlnGlyArg 90
: : ||||| |||||
330 ATCCTGCTTCGCAAGATCAG..... 349
91 TyrArgCysArgSerGlyLeuSerThrGlyTrp**GlnLeuSerLysLe 107
: : : ||||| |||||
350 .....GTAACACTCCCTGGCTGCCATCCAGTCCCT 381
107 leu.....GluLeuThrGlyProLysValLeuAlaCysSerL 120
: : : ||||| |||||
382 GATAGTCACTTCGATGCCGCGGACTTCTCCACGTCGTGTATGGTCA 431
120 euAlaLeuAspGlyAlaSer 126

```

```

:|||||
432 TGGCTCTGCGGCGTCC 451

```

seq_name: /cgn2_6/plodata/2/lna/5B_COMB.seq:US-08-431-333-3

seq_documentation_block:

Sequence 3, Application US/08431333

Patent No. 5965723

GENERAL INFORMATION:

APPLICANT: Shoyab, Mohammed

APPLICANT: Plozman, Gregory D.

TITLE OF INVENTION: EPITHELINS: NOVEL CYSTEINE-RICH GROWTH

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/431,333

FILING DATE: 27-APR-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/668,648

FILING DATE: 13-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: MISTOCK, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 5624-161-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)790-9090

TELEFAX: (212) 869-9741

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1779 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

TISSUE TYPE: kidney

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1779

US-08-431-333-3

```

alignment_scores:
Quality: 71.50 Length: 140
Ratio: 1.172 Gaps: 6
Percent Similarity: 43.571 Percent Identity: 23.571

```

alignment_block:

US-09-471-276-831 x US-08-431-333-3 ..

Align seg 1/1 to: US-08-431-333-3 from: 1 to: 1779

```

10 LeuTrpGlyValThrTrp.....GlyProValThrGluAlaAla 23
||||| :||||| |||||
80 CTGTGGCTCTGCTGCTGACCGCGAGGACGACGTAACCTCTCCCT 129
23 ephetyrGluThrGln.....**SerLeuTrpAlaGlu 35
: : : ||||| |||||
130 CCCCTCTGTGACAAATGCGCCACACACAGCAGGATCTGGTGCC 179

```



```

; APPLICATION NUMBER: US/08/464,517
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/963,308
; FILING DATE: 16-OCT-1992
; APPLICATION NUMBER: US 07/888,178
; FILING DATE: 26-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/701,514
; FILING DATE: 16-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Matthew P. Vincent
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MII-004C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-464-517-35
```

```

alignment_scores:
  Quality: 71.00      Length: 101
  Ratio: 1.315      Gaps: 4
  Percent Similarity: 53.465      Percent Identity: 26.733
```

alignment_block:

US-09-471-276-831 x US-08-464-517-35/rev ..

Align seq 1/1 to reverse of: US-08-464-517-35 from: 1 to: 1624

```

8 leuLeuLeuTrp.GlyValThrTrpGlyProValThrGluAlaIleP 24
   |||||
674 GTGCTTCTCTGAGAGGATAGATGTGATCTCCGACGAGAAACCATGCTGT 625
   |||||
24 heTyrgLurhrgIn**SerLeuTrpAlaGluSerGluHis**LeuLys 40
   |||||
624 TTCTGGGGGCTCTTCGGGGCGCGCGCGCGACGCG.....GTCCAG 584
   |||||
41 ThrLeuGlyGlnCysAspAlaAspValProGlyProGlyAspSerAr 57
   |||||
583 CGTCTAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534
   |||||
57 glauProAlaValGlnGluTrpGlyAlaGlnGluProValHisLeu... 72
   |||||
533 GCCAGCGCGCACGACGAGAAACCCGCTTCCTCCGCGCGCGCGCGCGTGA 484
   |||||
73 .AspSerPro.....AlaIleLysHisGlnPheLeu 83
   |||||
483 CAAGCCCGCGAGGTCTCTGCGCTGACGCGCGCGCACGCGATGCTTT 434
   |||||
84 ThrGlyAspThrGlnGlyArgTyArgCysArgSerGlyLeuSerThrG1 100
   |||||
433 CTAGCAAAATGCGCGCGAGGAGGAGCGCGCGATGACGATTAAGTCCG 384
   |||||
100 Y 100
383 C 383
```

seq_name: /cgn2_6/ptodata/2/1na/5b_COMB.seq:US-08-246-361A-35

seq_documentation_block:

```

; Sequence 35, Application US/08246361A
; Patent No. 5998362
; GENERAL INFORMATION:
; APPLICANT: BEACH, David H.
```

```

; TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/246,361A
; FILING DATE: 19-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/963,308
; FILING DATE: 16-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/888,178
; FILING DATE: 26-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/701,514
; FILING DATE: 16-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Matthew P. Vincent
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MII-004C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-246-361A-35
```

```

alignment_scores:
  Quality: 71.00      Length: 101
  Ratio: 1.315      Gaps: 4
  Percent Similarity: 53.465      Percent Identity: 26.733
```

alignment_block:

US-09-471-276-831 x US-08-246-361A-35/rev ..

Align seq 1/1 to reverse of: US-08-246-361A-35 from: 1 to: 1624

```

8 leuLeuLeuTrp.GlyValThrTrpGlyProValThrGluAlaIleP 24
   |||||
674 GTGCTTCTCTGAGAGGATAGATGTGATCTCCGACGAGAAACCATGCTGT 625
   |||||
24 heTyrgLurhrgIn**SerLeuTrpAlaGluSerGluHis**LeuLys 40
   |||||
624 TTCTGGGGGCTCTTCGGGGCGCGCGCGCGCGACGCG.....GTCCAG 584
   |||||
41 ThrLeuGlyGlnCysAspAlaAspValProGlyProGlyAspSerAr 57
   |||||
583 CGTCTAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534
   |||||
57 glauProAlaValGlnGluTrpGlyAlaGlnGluProValHisLeu... 72
   |||||
533 GCCAGCGCGCACGACGAGAAACCCGCTTCCTCCGCGCGCGCGCGTGA 484
   |||||
73 .AspSerPro.....AlaIleLysHisGlnPheLeu 83
   |||||
483 CAAGCCCGCGAGGTCTCTGCGCTGACGCGCGCGCACGCGATGCGCTTT 434
   |||||
```

```
84 ThrGlyAspThrGlnGlyArgCysArgSerGlyLeuSerThrG1 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
433 CTAGGAAATGCGCCCGGAGGAGGAGGCGAGTGAAGGATTAGTCCGG 384
100 Y 100
1
383 C 383

seq.name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-463-772-35

seq_documentation_block:
; Sequence 35, Application US/08463772
; Patent No. 6066501
; GENERAL INFORMATION:
; APPLICANT: BEACH, David H.
; TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(lexl)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,772
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/963,308
; FILING DATE: 16-OCT-1992
; APPLICATION NUMBER: US 07/888,178
; FILING DATE: 26-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/701,514
; FILING DATE: 16-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Matthew P. Vincent
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: M11-004C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-463-772-35

alignment_scores:
Quality: 71.00 Length: 101
Ratio: 1.315 Gaps: 4
Percent Similarity: 53.465 Percent Identity: 26.733

alignment_block:
US-09-471-276-831 x US-08-463-772-35/rev. ..
Align seg 1/1 to reverse of: US-08-463-772-35 from: 1 to: 1624
8 LeuLeuLeuTrp.GlyValThrTrpGlyProValThrGluAlaAlaIleP 24
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||
674 GTGCTTCTGTGAGGAGATGATGTATCTCGACGCAAAACCATCTGT 625
```

```
24 hetYrGluThrGln***SerLeuTrpAlaGluSerGluHis**LeuLys 40
||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
624 TTCTGGGGGCTCTTCGGCGCCCGCTGGGGCGACGCG.....GTGCAG 584
41 ThrLeuGlyGlnCysAspAlaAspValProGlyProProGlyAspSerAr 57
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
583 CGTCTAGGGCCGCCUUCAGGCCGCGGGCGACGGCTCCCGACGGGTTCCCGCCG 534
57 gluPProAlaValGlnGluTrpGlyAlaGlnGluProValHisLeu.... 72
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
533 GCCAGCGGCCGACGACAGAAACCCTTCCTCGCCCGCATCTGCTGA 484
73 ..AspSerPro.....AlaIleLysHisGlnPheLeuLeu 83
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
483 CAAGCCCGCCGAGGTGTCTGCGCTGAGCGGTGGCCACACCGATGCACCTTT 434
84 ThrGlyAspThrGlnGlyArgCysArgSerGlyLeuSerThrG1 100
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
433 CTAGGAAATGCGCCCGGAGGAGGAGGCGAGTGAAGGATTAGTCCGG 384
100 Y 100
1
383 C 383
```

```
seq.name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-05000-35

seq_documentation_block:
; Sequence 35, Application PC/TUS9305000
; GENERAL INFORMATION:
; APPLICANT: MITOTIX
; TITLE OF INVENTION: D-Type Cyclin and Uses Related Thereto
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05000
; FILING DATE: 19930525
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,178
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL91-02A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 616-861-9540
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US93-05000-35

alignment_scores:
Quality: 71.00 Length: 101
Ratio: 1.315 Gaps: 4
Percent Similarity: 53.465 Percent Identity: 26.733
```

alignment_block:
US-09-471-276-831 x PCT-US93-05000-35/rev ..

Align seg 1/1 to reverse of: PCT-US93-05000-35 from: 1 to: 1624

```
8 LeuLeuLeuTrp GlyValThrTrpGlyProValThrGluAlaIleP 24
   ::::::::::::::::::::
674 GTGCTTCTCTGAGGATGATGATGATGATGATGATGATGATGATG 625
   ::::::::::::::::::::
24 heTyrGluTrpGln***SerLeuTrpAlaGluSerGluHis***LeuLys 40
   ::::::::::::::::::::
624 TTCTGGGGGCTCTTGGGGGCTGGGGGCGACGCG.....GTGCAG 584
   ::::::::::::::::::::
41 ThrLeuGlyGlnCysAspAlaAspValProGlyProGlyAspSerAr 57
   ::::::::::::::::::::
583 CCGTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534
   ::::::::::::::::::::
57 GluLeuProAlaValGlnGluTrpGlyAlaGlnGluProValHisLeu... 72
   ::::::::::::::::::::
533 GCCACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 484
   ::::::::::::::::::::
73 ..AspSerPro.....AlaIleHisGlnPheLeuLeu 83
   ::::::::::::::::::::
483 CAAGCGCGCGCGAGTCTGCGCTGACCGTGGCCACACCGATGACGCTT 434
   ::::::::::::::::::::
84 ThrGlyAspThrGlnGlyArgTyrArgCysArgSerGlyLeuSerThrG 100
   ::::::::::::::::::::
433 CTACGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 384
   ::::::::::::::::::::
100 Y 100
   ::::::::::::::::::::
383 C 383
```

seq_name: /cgn2_6/pdata/2/lna/5B_COMB.seq:US-08-786-527A-1

seq_documentation_block:

```
; Sequence 1, Application US/08786527A
; Patent No. 5969210
; GENERAL INFORMATION:
; APPLICANT: Sparma and Montminy
; TITLE OF INVENTION: No. 5969210el Methods for the Characterization of Compounds w
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MCGREGOR & ADLER, P.C.
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,527A
; FILING DATE: January 21, 1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Benjamin A.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5848
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-777-2321
; TELEFAX: 713-777-6908
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 bp
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: other nucleic acid
```

```
;
;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; US-08-786-527A-1
```

alignment_scores:

Quality:	70.50	Length:	77
Ratio:	1.533	Gaps:	4
Percent Similarity:	59.740	Percent Identity:	32.468

alignment_block:

US-09-471-276-831 x US-08-786-527A-1 ..

Align seg 1/1 to: US-08-786-527A-1 from: 1 to: 403

```
50 ProGlyProProGlyAspSerArg.....LeuProAlaVal 61
   ::::::::::::::::::::
141 CCGCGCGCGCTCTTAAAGCAGCGCGCGCGCGCGCGCGCGCGCGCG 190
   ::::::::::::::::::::
61 GlnGluTrpGlyAlaGlnGluProValHisLeuAspSerProAlaI 77
   ::::::::::::::::::::
191 AACCAAGTGGGGGTGCTACAGAGCTATGACGCGCGCGCGCGCGCG 240
   ::::::::::::::::::::
77 eLysHisGlnPheLeuLeuThrGlyAspThrGlnGlyArgTyrArgCys 94
   ::::::::::::::::::::
241 CTCGCCACAGCCCTGTGTGTTCCCGGAGAGACAGTGCAGACTGTCAAC 290
   ::::::::::::::::::::
94 rGSerGlyLeuSerThrGlyTrp**GlnLeuSerLysLeuGluLeu 110
   ::::::::::::::::::::
291 GATCTGGGCGTGGCGCTG.....AGAGTCCGTGAGCTG 322
   ::::::::::::::::::::
111 ThrGlyProLysValLeuAlaCysSer 119
   ::::::::::::::::::::
323 CCAGCCGCTTAAGCCCTGCGCTGTACG 349
   ::::::::::::::::::::
```

seq_name: /cgn2_6/pdata/2/lna/5B_COMB.seq:US-08-667-939A-19

seq_documentation_block:

```
; Sequence 19, Application US/08667939A
; Patent No. 5998166
; GENERAL INFORMATION:
; APPLICANT: Luo, Shun
; TITLE OF INVENTION: CD16-II VARIANTS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/667,939A
; FILING DATE: 24-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/433,123
; FILING DATE: 03-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: LUO-2A
```

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 7..615
US-08-667-939A-19

alignment_scores:
Quality: 70.50 Length: 79
Ratio: 1.808 Gaps: 6
Percent Similarity: 49.367 Percent Identity: 34.177

alignment_block:
US-09-471-276-831 x US-08-667-939A-19 ..
Align seg 1/1 to: US-08-667-939A-19 from: 1 to: 630

43 GYVGINCYASPAALASPVALPROGLY.....PROPROGLY.ASPS 56
|||||.....|
126 GGACAGTGTGACTCTGAAGTCCACAGGAGCCTACTCCCTCAGACACATTT 175
56 erArgLeuProAlaValGInGluTrpGlyAlaGInGluProValHisLeu 72
|||.....|
176 CC.....ACACAGTGTTCACATGAG.....AGCCTC 204
73 ASPserProAlaIleLysHisGlnPheLeu.....LeuThrGlyAspThr 87
|||||.....|
205 ATCTCAAGCCAGGCGCTCGAGCTACTTATGACGCTGCACACAGTCGACGA 254
87 rGInGlyArGTYrArGcysArGserGlyLeuSerThr..... 99
|||||.....|
255 CAGTGGAGAGTACAGGTGCCAGACAACTCTCCACCTTCAGTGACCCGG 304
100 .....GlyTrp**GlnLeu 104
|||||
305 TGCAGCTAGAGTCCATATCGCGTGGCTGTGCTC 339

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-667-939A-17

seq_documentation_block:
; Sequence 17, Application US/08667939A
; Patent No. 5998166
; GENERAL INFORMATION:
; APPLICANT: LUD, Shun
; TITLE OF INVENTION: CD16-II VARIANTS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/667,939A
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/433,123
; FILING DATE: 03-MAY-1995
```

```
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: LUD-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 648 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..645
US-08-667-939A-17

alignment_scores:
Quality: 70.50 Length: 79
Ratio: 1.808 Gaps: 6
Percent Similarity: 49.367 Percent Identity: 34.177

alignment_block:
US-09-471-276-831 x US-08-667-939A-17 ..
Align seg 1/1 to: US-08-667-939A-17 from: 1 to: 648

43 GYVGINCYASPAALASPVALPROGLY.....PROPROGLY.ASPS 56
|||||.....|
69 GGACAGTGTGACTCTGAAGTCCACAGGAGCCTACTCCCTCAGACACATTT 118
56 erArgLeuProAlaValGInGluTrpGlyAlaGInGluProValHisLeu 72
|||.....|
119 CC.....ACACAGTGTTCACATGAG.....AGCCTC 147
73 ASPserProAlaIleLysHisGlnPheLeu.....LeuThrGlyAspThr 87
|||||.....|
148 ATCTCAAGCCAGGCGCTCGAGCTACTTATGACGCTGCACACAGTCGACGA 197
87 rGInGlyArGTYrArGcysArGserGlyLeuSerThr..... 99
|||||.....|
198 CAGTGGAGAGTACAGGTGCCAGACAACTCTCCACCTTCAGTGACCCGG 247
100 .....GlyTrp**GlnLeu 104
|||||
248 TGCAGCTAGAGTCCATATCGCGTGGCTGTGCTC 282

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-667-939A-12

seq_documentation_block:
; Sequence 12, Application US/08667939A
; Patent No. 5998166
; GENERAL INFORMATION:
; APPLICANT: LUD, Shun
; TITLE OF INVENTION: CD16-II VARIANTS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/667,939A
```

FILED DATE: 24-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/433,123
FILING DATE: 03-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: LUC-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 765 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-667-939A-12

alignment_scores:
Quality: 70.50 Length: 79
Ratio: 1.808 Gaps: 6
Percent Similarity: 49.367 Percent Identity: 34.177

alignment_block:

US-09-471-276-831 x US-08-667-939A-12 ..

Align seq 1/1 to: US-08-667-939A-12 from: 1 to: 765

43 GYGLINCYSAPALASPAVALPROGILY.....PROFROGILY ASPS 56
|||||.....||||| ||| ||| :||
120 GGACAGCTGACTCTGAAGTCCAGGAGACCTACTCCCTGAGGACAAATT 169
56 eTArGLeuProAlaValGInGluTrpGlyAlaGInGluProValHisLeu 72
|| :||| :||| |||
170 CC.....ACACAGTGTTCACATGTAG.....AGCCTC 198
73 AspSerProAlaIleLysHisGlnPheLeu.....LeuThrGlyAspThr 87
||| :||| ||| ||| |||
199 ATCTCAAGCCAGCGCTCGAGCTACTTCAATGACCGTCCACAGTCGACGA 248
87 rGInGIArGTyArGcysArGserGlyLeuSerThr..... 99
:||||:|||||:|||||:|||||
249 CAGTGGAGAGTACAGTGTCCAGACAAACCTCTCAACCTCAGTACCCGG 298
100GlyTrp**GlnLeu 104
299 TGCAGCTAGAAATCCATATCGGCTGCTGTTC 333

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-667-939A-15

seq_documentation_block:

Sequence 15, Application US/08667939A
Patent No. 5998166
GENERAL INFORMATION:
APPLICANT: LUD, Shun
TITLE OF INVENTION: CD16-II VARIANTS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/667,939A
FILING DATE: 24-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/433,123
FILING DATE: 03-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: LUC-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 765 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-667-939A-15

alignment_scores:
Quality: 70.50 Length: 79
Ratio: 1.808 Gaps: 6
Percent Similarity: 49.367 Percent Identity: 34.177

alignment_block:

US-09-471-276-831 x US-08-667-939A-15 ..

Align seq 1/1 to: US-08-667-939A-15 from: 1 to: 765

43 GYGLINCYSAPALASPAVALPROGILY.....PROFROGILY ASPS 56
|||||.....||||| ||| ||| :||
120 GGACAGTGTACTCTGAAGTCCAGGAGACCTACTCCCTGAGGACAAATT 169
56 eTArGLeuProAlaValGInGluTrpGlyAlaGInGluProValHisLeu 72
|| :||| :||| |||
170 CC.....ACACAGTGTTCACATGTAG.....AGCCTC 198
73 AspSerProAlaIleLysHisGlnPheLeu.....LeuThrGlyAspThr 87
||| :||| ||| ||| |||
199 ATCTCAAGCCAGCGCTCGAGCTACTTCAATGACCGTCCACAGTCGACGA 248
87 rGInGIArGTyArGcysArGserGlyLeuSerThr..... 99
:||||:|||||:|||||:|||||
249 CAGTGGAGAGTACAGTGTCCAGACAAACCTCTCAACCTCAGTACCCGG 298
100GlyTrp**GlnLeu 104
299 TGCAGCTAGAAATCCATATCGGCTGCTGTTC 333

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-667-939A-16

seq_documentation_block:

Sequence 16, Application US/08667939A
Patent No. 5998166
GENERAL INFORMATION:
APPLICANT: LUD, Shun
TITLE OF INVENTION: CD16-II VARIANTS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/667,939A
; FILING DATE: 24-JUN-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/433,123
; FILING DATE: 03-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: LEO-2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 765 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-667-939A-16

alignment_scores:
  Quality: 70.50      Length: 79
  Ratio: 1.808        Gaps: 6
  Percent Similarity: 49.367  Percent Identity: 34.177

alignment_block:
US-09-471-276-831 x US-08-667-939A-16 ..
Align seg 1/1 to: US-08-667-939A-16 from: 1 to: 765

43 GlyGlnCysAspAlaAspValProGly.....ProProGlyAsps 56
|||||
120 GGACAGTGTGACTGTGAGTCCAGGAGGAGCTACTCCCTGAGGACAAAT 169
56 eFArgLeuProAlaValGlnGluTrpGlyAlaGlnGluProValHisLeu 72
|||
170 CC.....ACACAGTGTGTTTCACATGAG.....AGCCTC 198
73 AspSerProAlaIleLeuHisGlnPheLeu.....LeuThrGlyAspTh 87
|||
199 ATCTCAAGCCAGGCGCTCGAGCTACTTCATTGAGCTGCGCACATCGACGA 248
87 rGlnGlyArgTyrArgCysArgSerGlyLeuSerThr..... 99
:::
249 CAGTGAGAGTACAGGTGCCAGACAAACCTCTCCACCCCTCAGTGACCCGG 298
100 .....GlyTrp**GlnLeu 104
299 TGCAGCTAGAGTTCATATTCGCTGCTGCTC 333

seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq:US-08-985-950-9

seq_documentation_block:
; Sequence 9, Application US/08985950
; Patent No. 6140076
; GENERAL INFORMATION:
; APPLICANT: Adema, Gosse Jan
; TITLE OF INVENTION: Isolated Mammalian Monocyte Cell Genes;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,950
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,279
; FILING DATE: 21-MARCH-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,181
; FILING DATE: 16-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,252
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0670K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1204
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 568 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 24..428
; NAME/KEY: mat-peptide
; LOCATION: 87..428
; US-08-985-950-9

alignment_scores:
  Quality: 69.00      Length: 83
  Ratio: 1.769        Gaps: 4
  Percent Similarity: 46.988  Percent Identity: 28.916

alignment_block:
US-09-471-276-831 x US-08-985-950-9 ..
Align seg 1/1 to: US-08-985-950-9 from: 1 to: 568

45 CysAspAlaAspValProGlyProProGlyAspSerArgLeuPro..... 59
|||||
155 TGTGACTTTCATGTGCGGGGCGGTTGGGCTTCAACATTCGCCCTGG 204
60 .....AlaValGlnGluT 64
205 AGAGGAGCATAGAGCCAGCACTACAAAGATGATATATGTTGCACTT 254
64 rPolyAlaGlnGluProVal.....HisLeuAspSerProAlaIleLeu 78
|||||
255 GTTCATCTGTGATGAGGAGCCAGATTCACATTCGACCA..... 293
78 SHISGlnPheLeuLeuThrGlyAspThrGlnGlyArgTyrArgCysArgS 95
294 .....GTAGTGAAGAAATGCCGGGCTTTATCGCTGCTCT 330
95 eGlyLeuSerThrGlyTrp**GlnLeuSerLysLeuGluLeu 110
|||||
331 ATTATAGCCGCCCTGATGCTGTGAGCAGCAGTACTTCCTGAGAGCTG 377

seq_name: /cgn2_6/plodata/2/lna/6B_COMB.seq:US-09-345-468-2

seq_documentation_block:
; Sequence 2, Application US/09345468
; Patent No. 6245527
; GENERAL INFORMATION:
```

```

: APPLICANT: Busfield, S.
: APPLICANT: Villaval, J.
: APPLICANT: Jandrot-Perrus, M.
: APPLICANT: Valinchenker, W.
: TITLE OF INVENTION: GLYCOPROTEIN VI AND USES THEREOF
: FILE REFERENCE: 7853-147
: CURRENT APPLICATION NUMBER: US/09/345,468
: CURRENT FILING DATE: 1999-06-30
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 1017
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-345-468-2

```

```

alignment_scores:
  Quality: 69.00      Length: 125
  Ratio: 1.131      Gaps: 6
  Percent Similarity: 48.800      Percent Identity: 27.200

```

alignment_block:

US-09-471-276-831 x US-09-345-468-2 ..

Align seg 1/1 to: US-09-345-468-2 from: 1 to: 1017

```

8 LeuLeuLeuTrpGlyValThrTrpGlyProValThrGlnAlaAlaIlePh 24
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
24 CTCTGCTTGGGCTGTCT...GGGGCTGTCCAGCGCAGAG.... 65
24 cTyrGluThrGln**SerLeuTrpAlaGluSerGluHis**LeuLys 41
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 .....TGGACCGCTCCSCAA.....G 81
41 hTrpGluGlnCysAspAlaAspValProGlyProPArgIleAspSerArg 57
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 CCCTCCCTCCAGCGCTGTCCAGCTGTCTGTCTGCGCCCTGGAGAACGAGT 131
58 LeuProAlaValGlnGluTrpGlyAlaGlnGluProValHisLeuAsp 74
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 GACCCCTCCGCTGCCAGGAGCTCCGGGCGTGGACCTGTACCGCTGGAGA 181
74 eT.....ProAla 76
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
182 AGCTGAGTTCCAGCAGGTACAGAGATCAGCAGTCTTCATCCGCGCC 231
77 IleLysHisGlnPheLeuLeuThrGlyAspTrpThrGlnGlyArgTyrArg 93
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 ATGAAAGAGAGTCTG.....GCTGGACGCTAACCGCTG 263
93 sArgSerGlyLeuSerThrGlyTrp**GlnLeuSerLysLeuLeuGlu 110
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
264 CTCTTCCACAGACGAGAGCTCTGTGCTCTGCCAGCGCAGCAGCTGGAGC 313
110 euThrGlyProLysValLeuAla 117
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
314 TCGTTCGACGAGGAGTTTTCGCC 336
seq_name: /cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-424-406-8
seq_documentation_block:
: Sequence 8, Application US/08424406
: Patent No. 5667997
: GENERAL INFORMATION:
: APPLICANT: Voelker, Toni Alois
: APPLICANT: Davies, Huw Maelor
: APPLICANT: Knutzen, Deborah S.
: TITLE OF INVENTION: Medium Chain Thioesterases in Plants
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Calgene, Inc.
: STREET: 1920 Fifth Street

```

```

: CITY: Davis
: STATE: CA
: COUNTRY: USA
: ZIP: 95616
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: Macintosh 7.0
: SOFTWARE: Microsoft Word 5.1(a)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/424,406
: FILING DATE: 26 April, 1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/10814
: FILING DATE: 29 October, 1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: USSN 07/968,971
: FILING DATE: 30 October, 1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Elizabeth Lassen
: REGISTRATION NUMBER: 31,845
: NAME: Donna E. Scherer
: REGISTRATION NUMBER: 34,719
: NAME: Carl J. Schwedler
: REGISTRATION NUMBER: 36,924
: REFERENCE/DOCKET NUMBER: CGNE 88-105
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (916) 753-6313
: TELEFAX: (916) 753-1510
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1474 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: US-08-424-406-8

```

```

alignment_scores:
  Quality: 69.00      Length: 113
  Ratio: 1.302      Gaps: 8
  Percent Similarity: 46.903      Percent Identity: 30.088

```

alignment_block:

US-09-471-276-831 x US-08-424-406-8 ..

Align seg 1/1 to: US-08-424-406-8 from: 1 to: 1474

```

33 AlAGluSerGluHis**LeuLysThrLeuGlyGlnCysAspAlaAspVa 49
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
790 TCAAAACGACGAGAGACTGTCAAAACT.....TCCAT 821
49 lProGlyProPArgIleAspSer.....ArgLeu.ProAlaVal 61
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
822 ACAGAGTTCCACGAGAGATAGTGCCTCTTTTGTGACACTCTCTGTCAAT 871
62 GlnGluTrpGlyAlaGlnGluProValHisLeuAspSerProAlaIle 78
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
872 GAA.....CAGAGTATCTGAAGT 891
78 sHisGlnPhe..LeuLeuThrLysAspThrGlnGlyArgTyrArgCysA 94
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
892 GCATAGTTTAAAGTGAAGACTGTGATTC.....ATTC 926
94 rGSerGlyLeuSerThrGlyTrp**GlnLeuSerLys..... 106
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
927 AAAAGGCTTAACCTCCGGGTGAATGACTTGGATGTCATCAATCAGACGTA 976
107 .....LeuLeuGluLeuThrGlyProLysVa 115
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

58 LeuProAlaValGlnGluTrpGlyAlaGlnIleuProVal.HisLeuAsp 74
   ||| ||| : : : : : ||| ||| : : : : :
167 GACCTCCGGTGGCAGGAGCTCCGGGCGTGGACCTGACCGCTGGAGA 216
   : : : : :
74 G.....ProAla 76
   : :
217 AGCTGACCTCCAGCAGCTACAGAGATGAGTCTTCATCTCCGCGC 266
   : :
77 IleHisGlnPheLeuLeuThrGlyAspThrGlnGlyArgTyrArg 93
   : : : : :
267 ATGAGAGAGAGTCTG.....GCTGACGCTACCCCTG 298
   : : : : :
93 sArgSerClyLeuSerThrGlyTrp**GlnLeuSerLysLeuGluL 110
   : : : : :
299 CTCTTACCAGAGGAGAGCCCTGCTGCTCCCTGCCAGGACGAGCTGAGC 348
   : : : : :
110 GUTHrGlyProLysValLeuAla 117
   : : : : :
349 TCGTTCACCGAGCTTTTCC 371

```

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-07-880-913-1

```

seq_documentation_block:
; Sequence 1, Application US/07880913
; Patent No. 5292651
; GENERAL INFORMATION:
; APPLICANT: GUTHRIE, ELLEN P.
; APPLICANT: VAN COTT, ELIZABETH M.
; APPLICANT: TAYLOR, CHRISTOPHER H.
; TITLE OF INVENTION: METHOD FOR PRODUCING THE NAEI
; TITLE OF INVENTION: RESTRICTION ENDONUCLEASE AND METHYLASE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/880,913
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, GREGORY D.
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: 41606
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 146..1099
; US-07-880-913-1

```

alignment_scores: Quality: 69.00 Length: 106
Ratio: 1.150 Gaps: 7

Percent Similarity: 56.604 Percent Identity: 33.019
alignment_block:
US-09-471-276-831 x US-07-880-913-1 ..

Align seq 1/1 to: US-07-880-913-1 from: 1 to: 3664

```

15 TrpGlyProValThrGluAlaIlePheThrGlyLeuThrGln**SerLe 31
   ||||| ||| ||||| : : : : : |||
2191 TGGGTCCTCCACACTCGGCGCGG.....TTTGTACTGTGCTCT.. 2232
   : : : : :
31 uTPAlaGluSerGluHis**LeuLysThrIleGlyGlnCysAsp.... 46
   ||||| : : : : : |||
2233 ..GCAGAACAAAGTTCGCCCCCTATTTCACCTGCGCTGAGCCGACCGGT 2278
   : : : : :
47 .....AlaAspValProGlyProProGlyAspSerArg 57
   : : : : :
2279 GCGGACCCACCGCTGGGAGAGCTTGAAGACCTTGAAGACCTTGAAGG 2328
   : : : : :
58 Leu.ProAlaValGlnGluTrpGlyAlaGlnIleuProValHisLeuAsp 74
   ||| : : : : : |||
2329 CTGGGAGAGTCCGAGAGAGTGGCGGCTCAG.....GCCAGACATCG 2372
   : : : : :
74 GProAlaIle.....LysHisGlnPheLeuLeuThrGly 85
   : : : : : |||||
2373 CACCAACATCTGCTGCTGCTGCTCCAGAAACATGCGGAGCTGACCTCGC 2422
   : : : : :
86 AspThrGlnGlyArgTyr.ArgCysArgSerGlyLeuSerThrClyTrp* 102
   ||| : : : : : |||
2423 CCGCTCTCGCGGAGAGCGGCGTGGGAGAGCTCGGTGTCAGCGCAAVGG 2472
   : : : : :
102 **GlnLeuSerLys 106
   : : : : :
2473 AGTGGTGGACCGC 2486

```

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-928-615-1

```

seq_documentation_block:
; Sequence 1, Application US/08928615
; Patent No. 565535
; GENERAL INFORMATION:
; APPLICANT: Chaux, Pascal
; APPLICANT: Stroobant, Vincent
; APPLICANT: Boon, Thierry
; APPLICANT: van der Bruggen, Pierre
; TITLE OF INVENTION: MAGR-3 PEPTIDES PRESENTED
; TITLE OF INVENTION: BY HLA CLASS II MOLECULES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,615
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-2441
; TELEFAX: 617-720-3500
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```



```

XX SQ Sequence 276 BP; 58 A; 82 C; 79 G; 57 T; 0 other;

alignment_scores:
    Quality: 372.00      Length: 81
    Ratio: 4.831        Gaps: 1
    Percent Similarity: 95.062    Percent Identity: 95.062

alignment_block:
US-09-471-276-831 x AAV88297 ..

Align seg 1/1 to: AAV88297 from: 1 to: 276

1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
|||||
37 ATGTCTCATGCTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 86
|||||
17 ova1ThrGluAlaAla1lePheTyrgluThrGln***SerLeuTrpAlaG 34
|||||
87 AGTGACGAGAAAGCAGCCATATTATGAGACGACCCAGCCCTGTGGCAG 136
|||||
34 luserGluHis**LeuLysThrLeuGlyGlnCysAspAlaAspValPro 50
|||||
137 AGTCCGATCATGCTGCTGAA.ACCCTTGCCCAATGTGACGCTGACGTCCA 185
|||||
51 GlyProGlyAspSerArgLeuProAlaValGlnGluTrpGly.AlaG 67
|||||
186 GGGCCGCTGAGACCTCCAGACTCCAGCTTCAGATGAGAAAGGGGTGGCC 235
|||||
67 lnc1uProValHisLeuAspSerProAla1leLysHisGln 80
|||||
236 AGGAGCCTGTGCACCTTGACTCACCCTCCATCAAGCAG 276

seq_name: /SID52/gcdata/geneseq/geneseqn/NA2001.DAT:MAC87257
seq_documentation_block:
ID AAC87257 standard; cdna: 1908 BP.

AC AAC87257;
XX
XX
XX 09-MAR-2001 (first entry)
XX
XX Mouse liver growth hormone-induced cdna clone 5, SPQ ID NO:7.
XX
XX Mouse; growth hormone; GH regulatable gene; liver pathology; hypertrophy;
XX hepatocellular lesion; hyperplasia; altered expression level; clone 5;
XX diagnostic marker; gigantism; acromegaly; diabetes; hepatotropic;
XX transgenic animal; drug screening; drug discovery; murine; ss.
XX
XX Mus sp.
XX
XX WO200066787-A2.
XX
XX 09-NOV-2000.
XX
XX 05-MAY-2000; 2000MO-US12366.
XX
XX 05-MAY-1999; 99US-0132663.
XX
XX (UYOH-) UNIV OHIO.
XX
XX Korchick JJ, Tiong J;
XX
XX WPT: 2001-007239/01.
XX P-PSDB: AAB48724, AAB48725, AAB48726, AAB48727, AAB48728, AAB48729,
XX AAB48730, AAB48731, AAB48732, AAB48733, AAB48734, AAB48735, AAB48736,
XX AAB48737, AAB48738, AAB48739.
XX
XX
XX Diagnosing abnormal levels of growth hormone activity in liver
XX expression level of hormone-regulatable liver genes, as diagnostic
XX markers of liver pathology

```

```

XX PS Claim 1: Page 51-52; 65pp; English.
XX
CC The invention relates to a method of diagnosing abnormal levels of
CC growth hormone (GH) activity in the liver, or predicting a change in the
CC condition of the liver in response to abnormal GH activity. The method
CC involves correlating the level of expression of certain specific genes
CC with the level of GH activity in the liver, or with an expected change
CC in the condition of the liver as the result of GH activity. Excessive GH
CC activity in the liver is thought to be deleterious to health, causing an
CC increase in liver size as a consequence of both hyperplasia and
CC hepatocellular hypertrophy, and hepatocellular lesions which progress with
CC age. Studies in transgenic mice which express high levels of bovine
CC growth hormone identified a number of genes whose expression in the
CC liver is altered by high GH levels. The genes which are upregulated are
CC those encoding alpha-fetoprotein, corticosteroid binding globulin,
CC fetuin, rab8-interacting protein, paraoxonase-3, cytochrome P45011A,
CC S-2 hydroxyacid oxidase, interferon alpha/beta receptor, growth hormone
CC receptor, proteasome 2-subunit, and coagulation factor V. Two novel
CC genes, clone 5 (AAC87257) and clone 45 (AAC87258) are also upregulated
CC in response to abnormally high GH levels. Conversely, expression of the
CC gene encoding 3-beta-hydroxysteroid dehydrogenase/delta-5-delta-4
CC isomerase appears to be downregulated. The invention also relates to
CC transgenic nonhuman mammals comprising a GH-induced transgene which
CC exhibit or have a propensity to develop a liver pathology; an assay for
CC drugs which inhibit the development of, or which treat a liver pathology,
CC comprising administering the drug to the transgenic animal; and
CC preventing or treating a liver pathology in a patient comprising
CC administering a drug which inhibits the expression of a GH-induced gene.
CC The method of the invention is used for diagnosing abnormal levels of GH
CC activity in the liver or predicting a change in the condition of the
CC liver in response to abnormal levels of GH activity. The GH-regulatable
CC liver genes and proteins are useful as diagnostic markers of liver
CC pathology. Assays for the expression of these genes is useful for the
CC diagnosis of liver pathologies associated with gigantism or acromegaly or
CC with diabetes, as other causative agents may act directly or indirectly
CC upon the same genes. The present sequence represents the novel mouse
CC liver cdna clone 5, expression of which is upregulated by abnormal GH
CC levels.
XX
XX SQ Sequence 1908 BP; 525 A; 475 C; 477 G; 431 T; 0 other;

alignment_scores:
    Quality: 139.50      Length: 138
    Ratio: 1.744        Gaps: 4
    Percent Similarity: 57.971    Percent Identity: 29.710

alignment_block:
US-09-471-276-831 x AAC87257 ..

Align seg 1/1 to: AAC87257 from: 1 to: 1908

1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
|||||
11 ATGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
|||||
17 ova1ThrGluAlaAla1lePheTyrgluThrGln***SerLeuTrpAlaG 34
|||||
61 AGGAATATCTTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 110
|||||
34 luserGluHis**LeuLysThrLeuGlyGlnCysAsp 46
|||||
111 AGCTTCAGTCTCTGAGACCTGGGCAACCTGACCTGCTGCTGCTGCA 160
|||||
47 AlaAspValProGlyProGlyAspSerArgLeuProAlaValGlnG 63
|||||
161 GTGATTTGCCG.....ACTAAGCTCTTCGAGCTGATTCAGCA 198
|||||
63 utrpGlyAlaGlnGluProValHisLeuAspSerProAla1leLysHisG 80
|||||
199 CGGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 248
|||||

```

```

80  InPheLeuLeu.....ThrGlyAspThrGlnGlyArgTyrArgCys 93
   :||||| ||| |||:|||||:||||| ||| ||| |||
249  CTTTTCCTGGGGGCGCATTAACAAGTAACAACAGCTGCATCTACCGCTGC 298
   94  ArgSerGlyLeuSer.....:|||||:||||| ThrG1 100
   ||| |||:|||||:|||||:|||||:|||||
299  AGATGTGGCGCTGGAGAACCCCTGTTGACATTTCACCTGCACACTGACAA 348
   100  yTTrp**GlnLeuSerLysLeuGlnLeuThrGlyProLysValLeuA 117
   |||||:|||||:|||||:|||||:||||| |||
349  CTGGACCATCTCAAGCAATGCTGTGGAGTGACACAGGAAAGAGCCCTTGC 398
   117  IAcYSerLeuAla 121
   |||||
399  CTCGGCCCTTGCGCT 412
seq_name: /SIDs2/gcsgdata/geneseq/geneseqn/NA2001.DAT:AA502105

seq_documentation_block:
ID  AA502105 standard; cDNA; 1477 BP.
XX
AC  AA502105;
XX
DT  18-JUL-2001 (first entry)
XX
DE  Human MANGO 511, variant #4 cDNA sequence.
XX
KW  Human; MANGO 511; transmembrane protein; diagnostic; asthma;
KW  Immunological disorder; arthritis; graft rejection; renal disorder;
KW  Acquired Immunodeficiency syndrome; inflammatory disorders; psoriasis;
KW  AIDS; embryonic disorder; brain; cerebral oedema; ischaemia; tumour;
KW  prostate; cerebrovascular disease; pituitary; Cushing's disease;
KW  neurodegenerative disease; Parkinson's disease; ss.
XX
OS  Homo sapiens.
XX
FH  Key Location/Qualifiers
FT  CDS 108..1007
FT  /'tag' a
FT  /product= "MANGO 511, variant #4"
XX
PN  WO200121631-A2.
XX
PD  29-MAR-2001.
XX
PF  20-SEP-2000; 2000WO-US25982.
XX
PR  20-SEP-1999; 99US-0399723.
XX
PA  (MILL-) MILLENNIUM PHARM INC.
XX
PI  Kirst SJ, Sharp JD, Fraser CC, Barnes T, Kingsbury G;
DR  WPI: 2001-211461/21.
DR  P-PSDB: AA001394.
XX
PT  New nucleic acid encoding INTERCEPT 307, MANGO 511, TANGO 351, TANGO
PT  361, TANGO 499 or TANGO 509 secreted or transmembrane protein, useful
PT  for the diagnosis and treatment of arthritis, psoriasis and Parkinson's
PT  disease -
XX
PS  Disclosure: Page 300-301; 362pp; English.
XX
XX
XX  The sequence represents the coding sequence of human MANGO 511
XX  variant #4 transmembrane protein. The nucleic acid and polypeptide
XX  sequences are useful for the diagnosis, prognosis and treatment of
XX  immunological disorders (e.g. arthritis, graft rejection and acquired
XX  immunodeficiency syndrome), inflammatory disorders (e.g. psoriasis and
XX  asthma), renal disorders, embryonic disorders, brain-related disorders
XX  (e.g. cerebral oedema), cerebrovascular diseases (e.g. ischaemia),
XX  tumours, prostate-related disorders, pituitary-related disorders (e.g.
XX  Cushing's disease) and neurodegenerative diseases (e.g. Parkinson's
XX  disease).

```

```

XX
SQ  Sequence 1477 BP; 393 A; 390 C; 372 G; 322 T; 0 other;

alignment_scores:
      Quality: 109.00      Length: 120
      Ratio: 1.627
      Percent Similarity: 55.833      Percent Identity: 34.167

alignment_block:
US-09-471-276-831 x AA502105 ..

Align seg 1/1 to: AA502105 from: 1 to: 1477

1  MetSerMetLeuValAlPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
   :|||:|||||:|||||:|||||:|||||:|||||
162  GTGAGCCCTCCATGCTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 211
   17  oValThrGluAlaAlaIlePheTyrGlnThrGln**SerLeuTrpAlaG 34
   | |||:|||||:|||||:|||||:|||||:|||||
212  CAAGACCCACAGTCGACGAGGAGGAAACCTCTCCAAAGCACCCCTGGGCTG 261
   34  IuSerGluHis**LeuLysThrLeuGlyGln..CysAspAlaAspValP 50
   || |||:|||||:|||||:|||||:|||||:|||||
262  AGCCAGGCTCTGTGATC...AGCCGGGGAACTCTGTGACCATCCGGTGT 308
   50  roGlyProProGlyAspSerArgLeuProAla ValGlnGlu..... 63
   ||| |||||:|||||:|||||:|||||:|||||:|||||
309  CAGGGACCCCTGGAGGCCCAAGAAATACCTGTGTTAAAGGAGGAAACCCC 358
   64  .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
   ||| |||:|||||:|||||:|||||:|||||:|||||
359  AGAACCTGGGACACACAGAACCCACATG.....GAACCCAGAGACA 399
   78  yHisGlnPheLeuLeuThrGlyAspThrGln.....GlyArgTyr 91
   || |||:|||||:|||||:|||||:|||||:|||||
400  AGGCGACATTTCTCCATCCATCCATGACAGACACACATGACGAGATAC 449
   92  ArgCysArgSerGlyLeuSerThrGlyTTrp**GlnLeuSerLysLeuLe 108
   |||||
450  CGCTGTACTACTACACACCCCTGCAGGCTGTGACAGCCAGCCAGCCCTT 499
   108  uGluLeu 110
   |||||
500  GGAGCTG 506

seq_name: /SIDs2/gcsgdata/geneseq/geneseqn/NA2000.DAT:AAA54610

seq_documentation_block:
ID  AAA54610 standard; cDNA; 956 BP.
XX
AC  AAA54610;
XX
DT  11-APR-2001 (first entry)
XX
DE  Leukocyte immunoglobulin like receptor coding sequence (LIR-9s1).
XX
KW  Leukocyte immunoglobulin like receptor; LIR; gene therapy;
KW  autoimmunity; autoimmune disorders; immune system; human; ds.
XX
OS  Homo sapiens.
XX
FH  Key Location/Qualifiers
FT  CDS 115..912
FT  /'tag' a
FT  /product= "Leukocyte immunoglobulin like receptor
XX
PN  WO200068383-A2.
XX
PD  16-NOV-2000.
XX
PF  12-MAY-2000; 2000WO-US13228.
XX

```



```

XX 12-MAY-1999: 99US-0310463.
XX PA (IMNV ) IMMUNEX CORP.
XX PI Cosman DJ, Anderson DM, Borges L;
XX NPI: 2000-687645/67.
XX DR P-PSDB: AAB04182.
XX
XX Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
XX PT treating autoimmune diseases and disease states with suppressed immune
XX PT function
XX
XX Claim 2: Page 109-110: 117pp: English.
XX
XX Host cells transfected with a vector which are capable of
XX CC expressing a leukocyte immunoglobulin like receptor (LIR), can be
XX CC used to produce the LIR polypeptide. LIR coding sequences may be
XX CC used in the gene therapy of disorders mediated directly or
XX CC indirectly by defective or insufficient amounts of any of the LIR
XX CC polypeptides. The LIR polypeptides can be used to treat autoimmune
XX CC diseases and disease states with suppressed immune function.
XX
XX Sequence 956 BP: 220 A: 291 C: 269 G: 176 T: 0 other:
XX
XX
XX alignment_scores:
XX      Quality: 108.00      Length: 120
XX      Ratio: 1.612      Gaps: 6
XX Percent Similarity: 55.833      Percent Identity: 34.167
XX
XX alignment_block:
XX US-09-471-276-831 x AA54610 ..
XX
XX Align seg 1/1 to: AA54610 from: 1 to: 956
XX
XX 1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyTr 17
XX      ::::: ::::: ||||| ::::: |||||
XX 169 GAGACCCCTGGCCCTCATGGTCTGCTGCGCGCTGAGCTGGGCC 218
XX 17 GAlThrGluAlaAlaIlePheTyrGluIuThrGln**SerLeuTrpAlaG 34
XX      | |||::: ::::: ::::: ||||| |||||
XX 219 CAGGAGCCACGTCGAGCGAGGAGCAACCTCTCAAAAGCACCCCTGGGCTG 268
XX
XX 34 IuSerGluHis**LeuIuSThrLeuGlyGln..CysAspAlaAspVal 50
XX      ::: ::::: ||||| ::::: |||||
XX 269 ACCGACGCTCTGTCATC...AGCGGGGGAACCTGTGACCATCCGGTCT 315
XX
XX 50 roGlyProProGlyAspSerArgLeuProAla.ValGlnIu..... 63
XX      ||| ||||| ::|||::: |||::: |||
XX 316 CAGGGGACCTTGGAGCCCGACAGATACCGTCTGGTTAAACAGCGAGGCC 365
XX
XX 64 .....TTPGlyAlaGlnIuProValHisLeuAspSerProAlaIleL 78
XX      ||| ::|||::: |||||::: ::||| |
XX 366 AGAACCCCTGGGACACACAGAACCCACTG.....GAGCCCAAGACA 406
XX
XX 78 yHisGlnPheLeuLeuThrGlyAspThrGln.....GlyArgTyr 91
XX      ||::||| ::::: ||||| ::||| |||||
XX 407 AGCCGAGATTCCTCCATCCCATCCATCGACAGACACCATGACGAGGAGTAC 456
XX
XX 92 ArgCysArgSerGlyLeuSerThrGlyTrp**GlnLeuSerIysLeuLe 108
XX      ||||| ::||| |||||::: |||
XX 457 CGCTGTACTACTACAGCCCTGACAGGCTGTCAGAGCCGACGCCACCCCT 506
XX
XX 108 uGluLeu 110
XX      |||||
XX 507 GGAGCTG 513
XX
XX seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2000.DAT:AA54608
XX
XX seq_documentation_block:
XX ID AA54608 standard; cDNA; 1016 BP.
XX

```

```

AC   AAA54608;
XX
XX   11-APR-2001 (first entry)
XX
XX   Leukocyte immunoglobulin like receptor coding sequence (LIR-9m1).
DE
XX
XX   Leukocyte immunoglobulin like receptor; LIR; gene therapy;
KM   autoimmunity; autoimmune disorders; immune system; human; ds.
XX
XX   Homo sapiens.
XX
XX   Key               Location/Qualifiers
XX   CDS               69..968
XX   FT                /*tag =
XX   FT                /product= Leukocyte immunoglobulin like receptor
XX
XX   MOZ00068383-A2.
XX
XX   16-NOV-2000.
XX
XX   12-MAY-2000; 2000MO-US13228.
XX
XX   12-MAY-1999; 99US-0310463.
XX
XX   (IMMV ) IMMUNEX CORP.
XX
XX   Cosman DJ, Anderson DM, Borges L;
XX
XX   WPI: 2000-687645/67.
XX   DR   P-PSDB: AAB04180.
XX
XX   Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
XX   PT   treating autoimmune diseases and disease states with suppressed immune
XX   PT   function
XX
XX   Claim 2; Page 104-105; 117pp: English.
XX
XX
XX   Host cells transformed with a vector which are capable of
XX   CC   expressing a leukocyte immunoglobulin like receptor (LIR), can be
XX   CC   used to produce the LIR polypeptide. LIR coding sequences may be
XX   CC   used in the gene therapy of disorders mediated directly or
XX   CC   indirectly by defective or insufficient amounts of any of the LIR
XX   CC   polypeptides. The LIR polypeptides can be used to treat autoimmune
XX   CC   diseases and disease states with suppressed immune function.
XX
XX   SQ   Sequence 1016 BP; 231 A; 310 C; 282 G; 193 T; 0 other;

alignment_scores:
    Quality: 108.00      length: 120
    Ratio: 1.612        Gaps: 6
    Percent Similarity: 55.833    Percent Identity: 34.167

alignment_block:
US-09-471-276-831 x AAA54608 ..

Align seg 1/1 to: AAA54608 from: 1 to: 1016

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
   ::::: ::::: ::::: ::::: ::::: ::::: :::::
123 GTGAGCGCCCTGCGCCATGCTCTGCTGCTGCGCTGCGGCTGAGTGGGCC 172
17 oValThrGluAlaIaIaIaIaPheTyrgLThrGln**SerLeuTrpAlaG 34
   I ::::: ::::: ::::: ::::: ::::: :::::
173 CAGGACCCACGTCGACGCGAAGGAACTCTCCAAAGCCACCTCTGGGCTG 222
34 InSerGluHis**LeuLysThrLeuGlyGln..CysAspAlaAspValP 50
   :: ::::: ::::: ::::: ::::: :::::
223 AGCGAAGCTCTGTATC...AGCGGGGGAACCTGTGAGACATCCGGTGT 269
50 roGlyProGlyAspSerAlaGluPProAla.ValGlnGlu..... 63
   :: ::::: ::::: ::::: ::::: :::::
270 CAGGGGACCCCTGAGGCCAGGAATATACCTGTGCTTGAAGAGGGGAACCC 319

```

```

64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
    ||| :|||:||||: :|||
320 AGAACCTGGGACACAGAACCCACTG.....GAGCCCAAGACA 360
78 yshHsGlnPheLeuLeuThrGlyAspThrGln.....GlyArgTyr 91
    ||| :|||: :|||: :|||:
361 AGCCGAGATTCTCCATCCCATGACAGACCAATCCATCCAGAGATAC 410
92 ArgCysArgSerGlyLeuSerThrGlyTyr**GlnLeuSerLysLeuLe 108
    |||||
411 CGCTGTACTACTACAGACCTGCAGAGCTGCTCAGAGCCAGACCCCT 460
108 uGluLeu 110
    |||||
461 GGAGCTG 467

seq_name: /SID2/gcgdata/geneseq/geneseqn/NA2001.DAT:AA502065
seq_documentation_block:
ID   AA502065 standard; cDNA; 1477 BP.
XX
AC   AA502065:
XX
DT   18-JUL-2001 (first entry)
XX
DE   Human MANCO 511 cDNA sequence.
XX
KW   Human; MANCO 511; transmembrane protein; diagnostic; asthma;
KW   immunological disorder; arthritis; graft rejection; renal disorder;
KW   acquired immunodeficiency syndrome; inflammatory disorders; psoriasis;
KW   AIDS; embryonic disorder; brain; cerebral oedema; ischaemia; tumour;
KW   prostate; cerebrovascular disease; pituitary; Cushing's disease;
KW   neurodegenerative disease; Parkinson's disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key      Location/Qualifiers
FT   CDS      108..1007
FT           /tag" a
FT           /product" "MANCO 511"
FT           /note" "The coding sequence (ORF) is specifically
FT           claimed in Claim 1"
FT   sig_peptide 108..230
FT           /*tag" b
FT   mat_peptide 231..1004
FT           /*tag" c
XX
FM   WO200121631-A2.
XX
PD   29-MAR-2001.
XX
PE   20-SEP-2000; 2000WO-US25982.
XX
PR   20-SEP-1999; 99US-0399723.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   K1rst SJ, Sharp JD, Fraser CC, Barnes T, Kingsbury G;
XX
PI   WPI: 2001-211461/21.
XX
PI   P-PSDB: AAU01330.
XX
PT   New nucleic acid encoding INTERCEPT 307, MANCO 511, TANGO 351, TANGO
PT   361, TANGO 499 or TANGO 509 secreted or transmembrane protein, useful
PT   for the diagnosis and treatment of arthritis, psoriasis and Parkinson's
PT   disease -
XX
PS   Claim 1; Fig 7; 362pp; English.
XX
CC   The sequence represents the coding sequence of human MANCO 511
CC   transmembrane protein. The nucleic acid and polypeptide sequences
CC   are useful for the diagnosis, prognosis and treatment of immunological

```

```

CC disorders (e.g. arthritis, graft rejection and acquired immunodeficiency
CC syndrome), inflammatory disorders (e.g. psoriasis and asthma), renal
CC disorders, embryonic disorders, brain-related disorders (e.g. cerebral
CC oedema), cerebrovascular diseases (e.g. ischaemia), tumours, prostate-
CC related disorders, pituitary-related disorders (e.g. Cushing's disease)
CC and neurodegenerative diseases (e.g. Parkinson's disease).
XX
SQ   Sequence 1477 BP; 392 A; 390 C; 373 G; 322 T; 0 other;

alignment_scores:
    Quality: 108.00      Length: 120
    Ratio: 1.612        Gaps: 6
    Percent Similarity: 55.833      Percent Identity: 34.167

alignment_block:
US-09-471-276-831 x AA502065 ..

Align seg 1/1 to: AA502065 from: 1 to: 1477

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyPyr 17
   :||| :|||:||||| :|||:|||||
162 GTGAGCCCTGCCCTCATGCTGTCTGCTCTGCCCTGGGCTGAGTCTGGGCC 211
17 oValThrGluAlaAlaIlePheTyrGluThrGln**SerLeuTrpAlaG 34
   | :|||:||||| :|||:|||||
212 CAGGAGCCACAGCTGCAGGACGAGAAACCTCTCCAAAGCCACCTCTGGGCTG 261
34 InsArgGluHis**LeuLysThrLeuGlyGln..CysAspAlaAspValP 50
   || :|||:||||| :|||:|||||
262 AGCCAGGCTCTGTGATC...AGCCGGGGGAACCTGTGAGACCAATCCGCTGT 308
50 roGlyProPogGlyAspSerArgLeuProAla.ValGlnGlu..... 63
   ||| :|||:||||| :|||:|||||
309 CAGGGAGCCCTGCAGGCCCAAGATACCTGCTGTTAAAGAGGAGGACCC 358
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
   ||| :|||:||||| :|||:|||||
359 AGAACCTGGGACACAGAACCCACTG.....GAGCCCAAGACA 399
78 yshHsGlnPheLeuLeuThrGlyAspThrGln.....GlyArgTyr 91
   ||| :|||: :|||: :|||:
400 AGGCGAGATTCTCCATCCCATGACAGACCAATCCATCCAGAGATAC 449
92 ArgCysArgSerGlyLeuSerThrGlyTyr**GlnLeuSerLysLeuLe 108
   |||||
450 CGCTGTACTACTACAGACCTGCAGAGCTGCTCAGAGCCAGACCCCT 499
108 uGluLeu 110
   |||||
500 GGAGCTG 506

seq_name: /SID2/gcgdata/geneseq/geneseqn/NA2001.DAT:AA502102
seq_documentation_block:
ID   AA502102 standard; cDNA; 1477 BP.
XX
AC   AA502102:
XX
DT   18-JUL-2001 (first entry)
XX
DE   Human MANCO 511, variant #1 cDNA sequence.
XX
KW   Human; MANCO 511; transmembrane protein; diagnostic; asthma;
KW   immunological disorder; arthritis; graft rejection; renal disorder;
KW   acquired immunodeficiency syndrome; inflammatory disorders; psoriasis;
KW   AIDS; embryonic disorder; brain; cerebral oedema; ischaemia; tumour;
KW   prostate; cerebrovascular disease; pituitary; Cushing's disease;
KW   neurodegenerative disease; Parkinson's disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key      Location/Qualifiers

```


Percent Similarity: 55.833 Percent Identity: 34.167

alignment_block:
US-09-471-276-831 x AAS02103

Align seg 1/1 to: AAS02103 from: 1 to: 1477

```

1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
   ::::: ::::: ::::: ::::: ::::: :::::
162 GTGAGCGCCCTGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 211
   ::::: ::::: ::::: ::::: ::::: :::::
17 oValThrGluAlaAlaIlePheTyrGluThrGln***SerLeuTrpAlaG 34
   ||| ::::: ::::: ::::: ::::: :::::
212 CAGGACCCAGCTGCGCAGGAGGAGAACCTCTCCAAAGCCACCCTCTGGG 261
   ||| ::::: ::::: ::::: :::::
34 InSerLeuHis***LeuLysThrLeuGlyGln..CysAspAlaAspValP 50
   ||| ::::: ::::: ::::: :::::
262 AGCCAGCGCTCTGTGATC...AGCCGGGGGAACTCTGTGACCATCCGGT 308
   ||| ::::: ::::: ::::: :::::
50 roGlyProPogIYAAspSerArgLeuProAla..ValGlnGlu..... 63
   ||| ::::: ::::: ::::: :::::
309 CAGGGACCCCTGGAGGCCCGCAGAAATACCGCTGCTTAAGAGGGAAGCC 358
   ||| ::::: ::::: ::::: :::::
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
   ||| ::::: ::::: ::::: :::::
359 AGAACCCCTGGACACAGAACCCACTG.....GAGCCCAAGAACCA 399
   ||| ::::: ::::: :::::
78 yshIsGlnPheLeuLeuThrGlyAspThrGln.....GlyArgTyr 91
   ||| ::::: ::::: :::::
400 AGGCCAGATTCTTCATCCATCCATGACAGACCATGACGAGGAGATAC 449
   ||| ::::: ::::: :::::
92 ArgCysArgSerGlyLeuSerThrGlyTrp***GlnLeuSerLysLeuLe 108
   ||| ::::: ::::: :::::
450 CGCTGTACTACTACAGCCCTCGAGGCTGTGTACAGCCCAAGCCACCCCT 499
   ||| ::::: :::::
108 uGluLeu 110
   ||| :::::
500 GCGAGCTG 506

seq_name: /SIND2/gcdata/geneseq/geneseqn/NA2001.DAT:AAS02104
seq_documentation_block:
ID AAS02104 standard; cDNA: 1477 BP.
XX AAS02104:
XX
XX 18-JUL-2001 (first entry)
XX
XX Human MANGO 511, variant #3 cDNA sequence.
XX
XX Human; MANGO 511; transmembrane protein; diagnostic; asthma;
XX immunological disorder; arthritis; graft rejection; renal disorder;
XX acquired immunodeficiency syndrome; inflammatory disorders; psoriasis;
XX AIDS; embryonic disorder; brain; cerebral oedema; ischaemia; tumour;
XX prostate; cerebrovascular disease; pituitary; Cushing's disease;
XX neurodegenerative disease; Parkinson's disease; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 108..1007
XX FT /*tag" a
XX FT /product" "MANGO 511, variant #3"
XX FT /transl_except" (pos:201..203, aa:Thr)
XX
XX MO200121631-A2.
XX
XX 29-MAR-2001.
XX
XX 20-SEP-2000; 2000MO-US25982.
XX
XX 20-SEP-1999; 99US-0399723.
XX

```

```

PA (MILL-) MILLENNIUM PHARM INC.
XX
XX Kirst SJ, Sharp JD, Fraser CC, Barnes T, Kingsbury G;
XX
XX WPI: 2001-211461/21.
XX
XX P-PSDB: AAU01393.
XX
XX New nucleic acid encoding INTERCEPT 307, MANGO 511, TANGO 351, TANGO
XX 361, TANGO 499 or TANGO 509 secreted or transmembrane protein, useful
XX for the diagnosis and treatment of arthritis, psoriasis and Parkinson's
XX disease -
XX
XX Disclosure: Page 298-299; 362pp; English.
XX
XX The sequence represents the coding sequence of human MANGO 511
XX variant #3 transmembrane protein. The nucleic acid and polypeptide
XX sequences are useful for the diagnosis, prognosis and treatment of
XX immunological disorders (e.g. arthritis, graft rejection and acquired
XX immunodeficiency syndrome), inflammatory disorders (e.g. psoriasis and
XX asthma), renal disorders, embryonic disorders, brain-related disorders
XX (e.g. cerebral oedema), cerebrovascular diseases (e.g. ischaemia),
XX tumours, prostate-related disorders, pituitary-related disorders (e.g.
XX Cushing's disease) and neurodegenerative diseases (e.g. Parkinson's
XX disease).
XX
XX Sequence 1477 BP; 392 A; 390 C; 373 G; 322 T; 0 other:

```

alignment_scores:
Quality: 108.00 Length: 120
Ratio: 1.612 Gaps: 6
Percent Similarity: 55.833 Percent Identity: 34.167

alignment_block:
US-09-471-276-831 x AAS02104

Align seg 1/1 to: AAS02104 from: 1 to: 1477

```

1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
   ::::: ::::: ::::: ::::: ::::: :::::
162 GTGAGCCCTGCCCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 211
   ::::: ::::: ::::: ::::: ::::: :::::
17 oValThrGluAlaAlaIlePheTyrGluThrGln***SerLeuTrpAlaG 34
   ||| ::::: ::::: ::::: :::::
212 CAGGACCCAGCTGCGCAGGAGGAGAACCTCTCCAAAGCCACCCTCTGGG 261
   ||| ::::: ::::: ::::: :::::
34 InSerLeuHis***LeuLysThrLeuGlyGln..CysAspAlaAspValP 50
   ||| ::::: ::::: :::::
262 AGCCAGCGCTCTGTGATC...AGCCGGGGGAACTCTGTGACCATCCGGT 308
   ||| ::::: ::::: :::::
50 roGlyProPogIYAAspSerArgLeuProAla..ValGlnGlu..... 63
   ||| ::::: ::::: :::::
309 CAGGGACCCCTGGAGGCCCGCAGAAATACCGCTGCTTAAGAGGGAAGCC 358
   ||| ::::: ::::: :::::
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
   ||| ::::: ::::: :::::
359 AGAACCCCTGGACACAGAACCCACTG.....GAGCCCAAGAACCA 399
   ||| ::::: ::::: :::::
78 yshIsGlnPheLeuLeuThrGlyAspThrGln.....GlyArgTyr 91
   ||| ::::: ::::: :::::
400 AGGCCAGATTCTTCATCCATCCATGACAGACCATGACGAGGAGATAC 449
   ||| ::::: ::::: :::::
92 ArgCysArgSerGlyLeuSerThrGlyTrp***GlnLeuSerLysLeuLe 108
   ||| ::::: ::::: :::::
450 CGCTGTACTACTACAGCCCTCGAGGCTGTGTACAGCCCAAGCCACCCCT 499
   ||| ::::: :::::
108 uGluLeu 110
   ||| :::::
500 GCGAGCTG 506

```

seq_name: /SIND2/gcdata/geneseq/geneseqn/NA1998.DAT:AAV44827

seq_documentation_block:

XX	ID	AAV44827 standard; DNA; 1550 BP.
XX	AC	AAV44827:
XX	DT	20-OCm-1998 (first entry)
XX	DE	FCR-IV coding sequence.
XX	XX	FC receptor-like protein; phagocytosis inducer; rheumatoid arthritis;
KM	KM	immune complex related disease; systemic lupus erythematosus; allergy;
KM	KM	haemolytic anaemia; thrombocytopenia; anaphylaxis; cancer; lymphoma;
KW	KW	leukaemia; infection; immunomodulator; viral entry inhibitor; therapy;
XX	XX	FCR-IV; ss.
OS	OS	Homo sapiens.
XX	XX	
FT	FT	Key
FT	FT	CDS
FT	FT	Location/Qualifiers
FT	FT	22..1440
FT	FT	/*tag= a
FT	FT	/product= FCR-IV
FT	FT	22..69
FT	FT	/*tag= b
FT	FT	70..1437
FT	FT	mat_peptide
FT	FT	/*tag= c
XX	XX	
PN	PN	MO9831806-A2.
XX	XX	
PD	PD	23-JUL-1998.
XX	XX	
PE	PE	20-JAN-1998; 98WO-US01184.
XX	XX	
PR	PR	18-JUN-1997; 97US-0049872.
PR	PR	21-JAN-1997; 97US-0034205.
XX	XX	
PA	PA	(HUMA-) HUMAN GENOME SCI INC.
XX	XX	
PI	PI	Gentz RL, Murphy M, Ni J, Olsen HS, Ruben SM;
XX	XX	
DR	DR	WPI: 1998-414105/35.
XX	XX	P-PSDB: AAM69234.
XX	XX	
PT	PT	Nucleic acid encoding Fc receptor-like polypeptides or their
PT	PT	fragments - and related vectors, transformed cells and antibodies,
PT	PT	useful for treating and diagnosing diseases of the haematopoietic
PT	PT	and immune systems
XX	XX	
PS	PS	Claim 2; Fig 4A; 141pp; English.
XX	XX	
CC	CC	This sequence encodes the Fc receptor-like IV protein (FCR-IV) of the
CC	CC	invention. Cells containing the DNA are used to express the recombinant
CC	CC	protein, and to screen for specific (ant)agonists. The proteins are used
CC	CC	to induce phagocytosis, and their (ant)agonists are used to treat immune
CC	CC	complex related diseases (e.g. rheumatoid arthritis, systemic lupus
CC	CC	erythematosus, haemolytic anaemia, thrombocytopenia, anaphylaxis,
CC	CC	allergy, colorectal or breast cancer, lymphoma, leukaemia, infection by
CC	CC	intracellular pathogens etc). The antagonists are also useful as
CC	CC	immunomodulators and inhibitors of viral (e.g. human immune deficiency or
CC	CC	dengue viruses) entry into cells. The proteins may also be used to screen
CC	CC	for specific binding agents, i.e. (ant)agonists, for raising antibodies
CC	CC	(Ab), and for identification of particular cells or tissues. The Ab can
CC	CC	be used therapeutically as antagonists; as assay reagents for diagnostic
CC	CC	determination of the levels of expression of the proteins and for
CC	CC	affinity purification of the proteins. The DNA and its fragments are
CC	CC	useful as hybridisation probes or primers for isolating related genes, in
CC	CC	situ hybridisation (chromosome mapping) and diagnostically to measure
CC	CC	mRNA expression.
XX	XX	
XX	XX	Sequence 1550 BP; 363 A; 506 C; 414 G; 267 T; 0 other;

```
alignment_scores:      Length: 121
      Quality: 104.50
      Ratio: 1.633
      Gaps: 7
```

```
Percent Similarity: 52.893      Percent Identity: 36.366
alignment_block:
US-09-471-276-831 x AAV44827 ..
Align seg 1/1 to: AAV44827 from: 1 to: 1550
```

```

2 SerLeuLeuValValPro...LeuLeuLeuTrpGlyValAlaTrpGly 16
   ::::: ::::: ::::: ::::: ::::: :::::
19 GCCATGATCCCCACCTTACAGGCTCTGCTCTGCCCTGGGCTGAGTGGG 68
   ::::: ::::: ::::: ::::: ::::: :::::
16 yProValThrGluAlaAlaIlePheTrpGluTrpGln**SerLeuTrpA 33
   ||| ||::: ::::: ::::: ::::: :::::
69 CCCAGTACCACATCGACGACGAGGCCCTCCCAAAACCAACCCCTGGG 118
   ::::: ::::: ::::: ::::: ::::: :::::
33 IaGluSerGluHis**LeuLysThrLeuLysGln...CysAspAlaAsp 48
   ||||| ::::: ||||| ::::: ||||| :::::
119 CTGAGCCAGGCTCTGTGATCAG...CTGGGGGAACCTGTGTGACCACTCG 164
   ::::: ::::: ::::: ::::: ::::: :::::
49 ValProGlyProProGlyAspSerArgLeuProAla.ValGlnGlu... 63
   ||| ||| ||||| ||| ::::: ||::: :::::
165 GTTCACAGGGGACCCCTGGAGGCTCGGAGATACCGCTGTGATTAAGAGGAA 214
   ::::: ::::: ::::: ::::: ::::: :::::
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAla 76
   ::::: ::::: ::::: ::::: ::::: :::::
215 GCCCAGCAGCCCTGGGACAGACAGAACCCACTG.....GAGCCCAAG 255
   ::::: ::::: ::::: ::::: ::::: :::::
77 IleLysHisGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArg 90
   ||| ::::: ||| ::::: ::::: ||| :::::
256 AACAGAGCCGCAATTTCTCCATTCCTCATACAGAGACATGACAGGAG 305
   ::::: ::::: ::::: ::::: ::::: :::::
90 gTyArgCysArgSerGlyLeuSerThrGlnGlyTrp**GlnLeuSerLysL 107
   ||||| ||| ::::: ||||| ||| :::::
306 ATACCCGCTGTTACTATGACGACCCGTGTAGGCTGTGTACAGCCCAAGTGACC 355
   ::::: ::::: ::::: ::::: ::::: :::::
107 euLeuGluLeu 110
   ||||| |||
356 CCTGTGAGACTG 366

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV69337

seq_documentation_block:

ID	AAV69337	standard; DNA; 1625 BP
AAV69337	standard; DNA; 1625 BP	

AC AAV69337;

DT 09-FEB-1999 (first entry)

Human LTR-pbm2 cDNA.

2025 LIR-pbm2: immunoregulator; leukocyte immunoglobulin-like receptor; KW
therapeutic; treatment; disorder; ITIM; MHC class I receptor; inhibitor KW
cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator; KW
negative signalling; autoimmune disease; suppressor; LIR; ss.

05 Homo sapiens.

FH	Key	Location/Qualifiers
FT	CDS	30.1376
FT		/*tag- a
XX		/product= "LIR-pbm2"
NN		

PD 29-OCT-1998.

PF 23-APR-1998; 98WO-US08244.

PR 24-APR-1997; 97US-0842248.

PA (IMMV) IMMUNEX CORP.,

PI Cosman DJ;

XX WP1; 1998-609990/51.
DR P-PSDB; AAW82551.
DR

PT Leukocyte immunoglobulin-like receptor, LIR, polypeptides - useful
PT e.g. for treating autoimmune diseases or disease states associated
PT with suppressed immune function

PS Example 9; Page 83-85; 112pp; English.

CC This sequence encodes a novel leukocyte immunoglobulin-like receptor
CC (LIR) polypeptide LIR-pbm2. This sequence can be administered
CC therapeutically to treat disorders associated with insufficient/defective
CC amounts of LIR polypeptide. LIR-p3g2 and certain other LIR family members
CC contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
CC (ITIMs), whilst other LIR family members lack ITIMs. By analogy with the
CC structure and function of known MHC Class I receptor molecules, LIRs
CC having ITIMs are inhibitory receptors mediating negative signalling,
CC whilst those lacking ITIMs are activatory receptors. Failure of a
CC receptor that mediates negative signalling could result in autoimmune
CC diseases, whilst failure of a receptor mediating activatory signalling
CC could result in suppressed immune function. They are also useful to
CC produce probes for detecting LIR nucleic acids or isolating LIR DNA from
CC other species.

SD Sequence 1625 BP; 399 A; 523 C; 425 G; 278 T; 0 other;

```
alignment_scores:
  Quality: 103.50
  Ratio: 1.617
  Percent Similarity: 52.893
  Length: 122
  Gaps: 7
  Percent Identity: 36.366
```

alignment_block:

US-09-471-276-831 x AAV69337 .

Align seg 1/1 to: AAV69337 from: 1 to: 1625

[illegible]

```
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV23273
seq_documentation_block:
```

ID AAV23273 standard; cDNA; 1625 BP.

AC AAV23273;

DT 17-JUL-1998 (first entry)

DE Human grp49 HM18 encoding cDNA.

KW Human; gpr49; HM18; HM43; immunoglobulin; immune response; mast cell;

OS Homo sapiens.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

PN W09809638-A1.

PD 12-MAR-1998.

PF 05-SEP-1997; 97WO-US15586

PR 06-SEP-1996; 96US-0025846.

PA (BGHM) BRIGHAM & WOMENS HOSPITAL

PI Arm JP, Austen KF, Castells MC, Katz HR;

DR WPI; 1998-193318/17.

DR P-PSDB; AAW53463.

PT Cell-surface member of immunoglobulin super-family, human gp43-
PT useful to treat undesired immune responses, especially mast
PT cell-related diseases
XX
PS Claim 5; Fig 1A-B; 62pp; English.

CC The present sequence encodes human

CC The present sequence encodes human gp49 polypeptide 1m18. The present
CC invention also describes: (1) a fusion polypeptide comprising a human
CC gp49 component and FcER1 or a detectable marker; (2) a recombinant
CC nucleic acid encoding human gp49 or a human gp49-FcER1 fusion as above;
CC (3) a cell or vector comprising the recombinant nucleic acid as in (2);
CC and (4) an antibody which selectively binds to gp49. Mammalian gp49 or
CC its related DNA can be used to treat an undesired immune response,
CC especially a mast cell-related disease.

SQ Sequence 1625 BP; 400 A; 519 C; 429 G; 277 T; 0 other;

alignment_scores:	
Quality:	103.50
Ratio:	1.617
Percent Similarity:	52.893
	Percent Identity: 36.364

alignment_block:

US-09-471-276-831 x AAV23273 . .

Align seg 1/1 to: AAV23273 from: 1 to: 1625

```

2 SerMetLeuValValPhe.....LeuLeuLeuThrPclValThrTrpCl 16
   ::::: ::::: ::::: ::::: :::::
43 GCGATGATCCCGCCACCTTCAGCTCTGCTGCCGCGAGATCTGGC 92
   ::::: ::::: ::::: ::::: :::::
16 yProValThrGluAlaAlaIlePheTyrglThrGln***SerLeuTrpA 33
   |||| ::::: ::::: ::::: :::::
93 CCCCGAGGCCCATGCGAGCGAGGCCCTCCCAACCCACCTCTGGC 142
   |||| ::::: ::::: ::::: :::::
33 IagLusecluhis**LeuTyThrIleuLgln...GysAraAlaAsp 48
   |||| ::::: ::::: ::::: :::::
143 CTGACCACGCTCTGTGTGAC...CTGGGGGAACTCTGAGCCATCTG 188

```

```

49 ValProGlyProProGlyAspSerArgLeuProAla.ValGInGlu.... 63
   ||| ||| ||||| ||| ::|||::: ::|||
189 GTGTCAAGGACCCCTGGAGCTCGGAGTACCTCTCGATAAAGAGGAA 238
64 .....TTPGlyAlaGInGluProValHisLeuAspSerProAla 76
   ||| ||| |||::|||::: ::|||
239 GCCCAGCACCCTGGAGACAGACAGACCACCTG.....GAGCCCAAG 279
77 IleLeuHisGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArg 90
   ||| ::||| ||| ::||| ||| |||||
280 AACAGGCCGAGATTCTCCATCCATCCATGACAGAGCATATGACAGGAG 329
90 gTyArgCysArgSerGlyLeuSerThrGlyTyrP**GlnLeuSerLysL 107
   ||||| ||| ::||| ||||| ||| |||
330 ATACCCGCTGTACTATGCGAGCCCTGTAGGCTGTGTACAGCCAGTGACC 379
107 euleuGluLeu 110
   ||||| |||
380 CCCTGGAGCTG 390
seq_name: /SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA54600

```

```

seq_documentation_block:
ID   AAA54600 standard; cDNA; 1625 BP.
XX
AC   AAA54600;
XX
DT   11-APR-2001 (first entry)
XX
DE   Leukocyte immunoglobulin like receptor coding sequence (phm2).
XX
KW   Leukocyte immunoglobulin like receptor; LIR; gene therapy;
KW   autoimmunity; autoimmune disorders; immune system; human; ds.
XX
OS   Homo sapiens.
XX
FH   Location/Qualifiers
FT   CDS
FT       30..1376
FT       /tag="a
FT       /product="Leukocyte immunoglobulin like receptor
XX
PN   WO200068383-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13228.
XX
PR   12-MAY-1999; 99US-0310463.
XX
PA   (IMV) IMMUNEX CORP.
XX
PI   Cosman DJ, Anderson DM, Borges L;
XX
DR   WPI: 2000-687645/67.
XX
DR   P-PSDB; AAB04176.
XX
PT   Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
PT   treating autoimmune diseases and disease states with suppressed immune
PT   function
XX
PS   Example 9; Page 88-90; 117pp: English.
XX
CC   Host cells transformed with a vector which are capable of
CC   expressing a leukocyte immunoglobulin like receptor (LIR), can be
CC   used to produce the LIR polypeptide. LIR coding sequences may be
CC   used in the gene therapy of disorders mediated directly or
CC   indirectly by defective or insufficient amounts of any of the LIR
CC   polypeptides. The LIR polypeptides can be used to treat autoimmune
CC   diseases and disease states with suppressed immune function.
XX
SQ   Sequence 1625 BP; 399 A; 523 C; 425 G; 278 T; 0 other;

```

```

alignment_scores:
Quality: 103.50      Length: 121
Ratio: 1.617        Gaps:
Percent Similarity: 52.893      Percent Identity: 36.364
alignment_block:
US-09-471-276-831 x AAA54600

```

Align seg 1/1 to: AAA54600 from: 1 to: 1625

```

2 SerMetLeuValValPhe.....LeuLeuThrProGlyValThrTrpG1 16
   ::|||::: ::||| ||||| |||::|||
27 GCCATGATCCCCACCTTCACGCGCTCTGCTGCCCTGGAGTGTGGG 76
16 yProValThrGluAlaAlaIlePheTyrGluThrGln**SerLeuTrpA 33
   |||| |||::: ::||| ||||| |||
77 CCCAGAGACCCACATGCGAGCGAGGCCCTCCCAAAACCCACCTCTGGG 126
33 IagIuSerGluHis**LeuLysThrLeuGlnGln...CysAspAlaasp 48
   ||||| ||| ::||| ||||| |||
127 CTGAGCCAGGCTCTGTGATCAG...CTGGGGAGACCTCTGTGACCATCTG 172
49 ValProGlyProProGlyAspSerArgLeuProAla.ValGInGlu.... 63
   ||| ||| ||||| ||| ::|||::: ::|||
173 GTGTCAAGGACCCCTGGAGCTCGGAGTACCTCTCGATAAAGAGGAA 222
64 .....TTPGlyAlaGInGluProValHisLeuAspSerProAla 76
   ||| ||| |||::|||::: ::|||
223 GCCCAGCACCCTGGAGACAGACAGACCACCTG.....GAGCCCAAG 263
77 IleLeuHisGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArg 90
   ||| ::||| ||| ::||| ||| |||||
264 AACAGGCCGAGATTCTCCATCCATCCATGACAGAGCATATGACAGGAG 313
90 gTyArgCysArgSerGlyLeuSerThrGlyTyrP**GlnLeuSerLysL 107
   ||||| ||| ::||| ||||| ||| |||
314 ATACCCGCTGTACTATGCGAGCCCTGTAGGCTGTGTACAGCCAGTGACC 363
107 euleuGluLeu 110
   ||||| |||
364 CCCTGGAGCTG 374
seq_name: /SID2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV69329
seq_documentation_block:
ID   AAV69329 standard; DNA; 2777 BP.
XX
AC   AAV69329;
XX
DT   09-FEB-1999 (first entry)
XX
DE   Human LIR-18A3 cDNA.
XX
KW   LIR-18A3; immunoregulator; leukocyte immunoglobulin-like receptor; LIR;
KW   therapeutic; treatment; disorder; ITIM; MHC class I receptor; inhibitor;
KW   cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
KW   negative signalling; autoimmune disease; suppressor; ss.
XX
OS   Homo sapiens.
XX
FH   Location/Qualifiers
FT   CDS
FT       168..2126
FT       /tag="a
FT       /product="LIR-18A3"
XX
PN   WO9848017-A1.
XX
PD   29-OCT-1998.
XX
PF   23-APR-1998; 98WO-US08244.
XX
PR   24-APR-1997; 97US-0842248.
XX

```



```

168 ATGACCCCATCTCTCAGGCTGATCTCTCGGGCTGAGTCTGGGCC 217
17 oValThrGluAlaAlaIlePheTyrGluThrGln***SerLeuTrpAlaG 34
218 CAGGACCCACAGTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 267
34 IuSerGluHis***LeuLysThrLeuGly..GlnCysAspAlaAspValP 50
268 AACGAGGCTCTGTGATC...ACCGAGGGGAGTCTCTGAGCCCTGAGTGT 314
50 roGlyProProGlyAspSerArgLeuProAlaVal..GlnGlu..... 63
315 CAGGGGGGCGCAGAGGAGCCAGGAGTACCGCTGTATATAGAGAAAGAAC 364
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
365 AGCAGCTGTGATTACACGGATCCACAGAGAGCTT.....GTGAGA 405
78 yshsGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArgTyr 91
406 AGGGCCAGTCCCGCCATCCATCCATCCATCCGAGACATGACAGGGCGTAT 455
92 ArgCysArgSerGlyLeuSerThr...GlyTrp***GlnLeuSerLysI 107
456 CGCTGTACTATGTAGCAGACTGAGCGCCCTCAGAGAGAGTGAACC 505
107 uLeuGluLeuThr..... 111
506 CTGGAGCTGTGTGTGACAGAGCCCTACATCAAAACCCCTCTCAGCCC 555
112 .....GlyProLysValLeuAlaCys... 118
556 AGCCAGCCCGCGTGTGAATCAGAGGAGATGATCCCTCCAGTGTGAC 605
119 ...SerLeuAlaLeuAspGlyAlaSer 126
606 TCACAGGTGTGATTTGATGCTTCAGT 632
seq_name: /sids2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV38608
seq_documentation_block:
ID AAV38608 standard; cDNA: 2790 BP.
XX
AC AAV38608;
XX
DT 23-SEP-1998 (first entry)
XX
DE Human monocyte gene designated KLM67.
XX
KW Human; type I transmembrane protein; immunoglobulin-like domain;
KW FcR3; activated monocyte; YE01; KTE03; control; development;
KW differentiation; mammalian immune system; treatment; cancerous condition;
KW degenerative condition; autoimmune response; transplantation rejection;
KW graft versus host disease; inflammatory condition; diagnosis;
KW drug screening; ss.
XX
OS Homo sapiens.
XX
FH Key - Location/Qualifiers
FT CDS 177..2135
FT /tag= a
FT /product= KLM67
XX
XX WO9824906-A2.
XX
PD 11-JUN-1998.
XX
PF 05-DEC-1997; 97WO-US21101.
XX
XX 21-MAR-1997; 97US-0041279.
XX 06-DEC-1996; 96US-0032252.
XX 09-DEC-1996; 96US-0762187.
XX 16-DEC-1996; 96US-0033181.

```

```

XX
PA (SCHE ) SCHERING CORP.
XX
PI Adema GJ, Gorman DM, Lanier LL, McClanahan TK, Meynard L,
PI Phillips JH, Zurawski G, Zurawski SM;
XX
DR WPI: 1998-333325/29.
XX P-PSDB: AAM62782.
XX
PT New isolated activated monocyte cell gene(s) - used to develop
PT products for treating e.g. cancer, degenerative conditions,
PT autoimmune responses, transplant rejection or inflammatory
PT conditions
XX
PS Disclosure: Pages 95-98: 104pp: English.
XX
CC AAV38992-96 and AAV38608 represent human monocyte genes, collectively
CC designated KTE03. The genes are found in activated monocytes. The
CC specification also describes other proteins encoded by activated
CC monocytes, which are designated FcR3 and YE01. The genes function in
CC controlling development, differentiation, and/or physiology of the
CC mammalian immune system. The products can be used for treating abnormal
CC proliferation, regeneration, degeneration or atrophy. They can be used
CC for treating e.g. cancerous conditions, degenerative conditions,
CC autoimmune responses, transplantation rejection, graft versus host
CC disease, or inflammatory conditions. The products can also be used for
CC detection, diagnosis and drug screening.
XX
SQ Sequence 2790 BP; 665 A; 847 C; 752 G; 525 T; 1 other:

```

```

alignment_scores:
  Quality: 98.50      Length: 121
  Ratio: 1.449        Gaps: 7
  Percent Similarity: 56.198   Percent Identity: 33.884

```

alignment_block:

US-09-471-276-831 x AAV38608

Align seg 1/1 to: AAV38608 from: 1 to: 2790

```

1 MetSerMetLeuValAlaPheLeuLeuTrpGlyValThrTrpGlyPr 17
111:..... 111:..... 111:..... 111:
177 ATGACCCCATCTCTCAGGCTGATCTCTCGGGCTGAGTCTGGGCC 226
17 oValThrGluAlaAlaIlePheTyrGluThrGln***SerLeuTrpAlaG 34
227 CCGGACCCACAGTGCAGGAGGAGCACCCTCCCAAGCCACCTCTGGGCTG 276
34 IuSerGluHis***LeuLysThrLeuGly..GlnCysAspAlaAspValP 50
277 AACGAGGCTCTGTGATC...ACCGAGGGGAGTCTCTGAGCCCTGAGTGT 323
50 roGlyProProGlyAspSerArgLeuProAlaVal..GlnGlu..... 63
324 CAGGGGGGCGCAGAGGAGCCAGAGTACCGCTGTATATAGAGAAAGAAC 373
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
374 AGCAGCTGTGATTACACGGATCCACAGAGAGCTT.....GTGAGA 414
78 yshsGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArgTyr 91
415 AGGGCCAGTCCCGCCATCCATCCATCCGAGACATGACAGGGCGGTAT 464
92 ArgCysArgSerGlyLeuSerThr...GlyTrp***GlnLeuSerLysI 107
465 CGCTGTACTATGTAGCAGACTGACAGGCCCTCAGAGAGAGTGAACC 514
107 uLeuGluLeu 110
515 CTGGAGCTG 524

```


XX Host cells transformed with a vector which are capable of
CC expressing a leukocyte immunoglobulin like receptor (LIR), can be
CC used to produce the LIR polypeptide. LIR coding sequences may be
CC used in the gene therapy of disorders mediated directly or
CC indirectly by defective or insufficient amounts of any of the LIR
CC polypeptides. The LIR polypeptides can be used to treat autoimmune
CC diseases and disease states with suppressed immune function.
XX
XX Sequence 2922 BP; 683 A; 882 C; 789 G; 568 T; 0 other:

alignment_scores:
Quality: 98.50 Length: 121
Ratio: 1.449 Gaps: 7
Percent Similarity: 56.198 Percent Identity: 33.884

alignment_block:

US-09-471-276-831 x AA54591 ..

Align seg 1/1 to: AA54591 from: 1 to: 2922

```
1 MetSerMetLeuValValPheLeuLeuThrPGLyValThrTrpGlyPr 17
||||| : : : : : ||||| : : : : : |||||
310 ATGACCCCTCCTCCTCCTGATCTCTCGGCTGAGTCTGGGCC 359
17 oValThrGluAlaIlePheTyrGluThrGln**SerLeuTrpAlaG 34
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
360 CCGGACCCACGTCGACGAGGACCTCCCAAGCCACCCTCTGGGCTG 409
34 lUserGluHis**LeuLysThrLeuGly..GlnCysAspAlaAspVal 50
|| : : : : : ||| : : : : : ||| : : : : :
410 AACGAGGCTCTGTGATC...ACCCAGGGAGTCTGTGACCCCTGAGTGT 456
50 roGlyProGlyAspSerArgLeuProAlaVal..GlnGlu..... 63
||| : : : : : ||| : : : : : ||| : : : : :
457 CAGGGGGGCGCAGAGCCAGGAGTACCGTCTATATAGAAAGAAAC 506
64 .....TTPGLYAlaGlnGluProValHisLeuAspSerProAlaIleL 78
||| : : : : : ||| : : : : : ||| : : : : :
507 AGCACCTCGATTACACGAGTCCACAGAGACTT.....GTGAAGA 547
78 yshHisGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArgTyr 91
||| : : : : : ||| : : : : : ||| : : : : :
548 AGGGCCAGTCCCATCCATCCATCATTGAGACATGACAGGGCGTAT 597
92 ArgCysArgSerGlyLeuSerThr...GlyTyr**GlnLeuSerLysLe 107
||| : : : : : ||| : : : : : ||| : : : : :
598 CGCTGTACTATGTGTAGCAGACACTGAGCGCGCTCAGAGAGAGTACCC 647
107 uLeuGlnLeu 110
||||| : : : : :
648 CCGTGGAGCTG 657
```

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:AA54611

seq_documentation_block:

ID AA54611 standard; cDNA: 997 BP.

XX AAA54611;

XX 11-APR-2001 (first entry)

DE Leukocyte immunoglobulin like receptor coding sequence (LIR-9s2).

XX Leukocyte immunoglobulin like receptor; LIR; gene therapy;

KW autoimmunity; autoimmune disorders; immune system; human; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 73..834 /tag= a

/product= Leukocyte immunoglobulin like receptor

FT
XX WO200068383-A2.
XX 16-NOV-2000.
XX 12-MAY-2000; 2000WO-US13228.
XX 12-MAY-1999; 99US-0310463.
XX (IMNV) IMMUNEX CORP.
XX Cosman DJ, Anderson DM, Borges L;
XX WPI: 2000-687645/67.
XX P-PSDB: AAB04183.

XX Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
PT treating autoimmune diseases and disease states with suppressed immune
PT function

PS Claim 2: Page 112-113; 117pp: English.

XX Host cells transformed with a vector which are capable of
CC expressing a leukocyte immunoglobulin like receptor (LIR), can be
CC used to produce the LIR polypeptide. LIR coding sequences may be
CC used in the gene therapy of disorders mediated directly or
CC indirectly by defective or insufficient amounts of any of the LIR
CC polypeptides. The LIR polypeptides can be used to treat autoimmune
CC diseases and disease states with suppressed immune function.
XX

SQ Sequence 997 BP; 241 A; 292 C; 281 G; 183 T; 0 other:

alignment_scores:
Quality: 90.50 Length: 75
Ratio: 2.207 Gaps: 4
Percent Similarity: 54.667 Percent Identity: 37.333

alignment_block:

US-09-471-276-831 x AA54611 ..

Align seg 1/1 to: AA54611 from: 1 to: 997

```
45 CysAspAlaAspValProGlyProGlyAspSerArgLeuProAla..Y 61
||||| : : : : : ||||| : : : : : |||||
222 TGTGACCATCCGCTGTACAGGACCTGGAGGCCGACGAAATACCTCTGG 271
61 aGlnGlu.....TTPGLYAlaGlnGluProValHisLeu 72
||| : : : : : ||| : : : : : ||| : : : : :
272 TTAAAGAGGAGAGCCAGAACCTGGGACACACAGAACCCACTG..... 315
73 AspSerProAlaIleLysHisGlnPheLeuLeuThrGlyAspThrGln.. 88
||| : : : : : ||| : : : : : ||| : : : : :
316 ..GAGCCCAAGAAAGGACGATCTCCATCCATCCATGACAGACA 362
89 .....GlyArgTyrArgCysArgSerGlyLeuSerThrGlyTyr**G 103
||||| : : : : : ||||| : : : : : |||||
363 CCATGCGAGGAGATACCGCGTACTACTACACGCTCGAGGCTGTGAG 412
103 lInLeuSerLysLeuLeuGlnLeu 110
||| : : : : : ||| : : : : : ||| : : : : :
413 AGCCCAAGCAGCCCTGTGAGCTG 435
```

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:AA54609

seq_documentation_block:

ID AA54609 standard; cDNA: 1007 BP.

XX AAA54609;

XX 11-APR-2001 (first entry)

XX

```
DE Leukocyte immunoglobulin like receptor coding sequence (LIR-9m2).
XX
KW Leukocyte immunoglobulin like receptor; LIR; gene therapy;
KW autoimmunity; autoimmune disorders; immune system; human; ds.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT CDS 95..958
FT /tag.. a
FT /product.. Leukocyte immunoglobulin like receptor
XX
PN WO200068383-A2.
XX
PD 16-NOV-2000.
XX
PP 12-MAY-2000; 2000WO-US13228.
XX
PR 12-MAY-1999; 99US-0310463.
XX
PA (IMMV ) IMMUNEX CORP.
XX
PI Cosman DJ, Anderson DM, Borges L;
XX
DR WPI: 2000-687645/67.
DR P-PSDB: AAB04181.
XX
XX Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
PT treating autoimmune diseases and disease states with suppressed immune
PT function
XX
XX Claim 2: Page 107-108; 117pp; English.
XX
XX Host cells transformed with a vector which are capable of
XX expressing a leukocyte immunoglobulin like receptor (LIR), can be
XX used to produce the LIR polypeptide. LIR coding sequences may be
XX used in the gene therapy of disorders mediated directly or
XX indirectly by defective or insufficient amounts of any of the LIR
XX polypeptides. The LIR polypeptides can be used to treat autoimmune
XX diseases and disease states with suppressed immune function.
XX
XX Sequence 1007 BP; 230 A; 303 C; 276 G; 198 T; 0 other;

alignment_scores:
Quality: 90.50 Length: 75
Ratio: 2.207 Gaps: 4
Percent Similarity: 54.667 Percent Identity: 37.333

alignment_block:
US-09-471-276-831 x AAA54609 ..

Align seg 1/1 to: AAA54609 from: 1 to: 1007

45 CysAspAlaAspValProGlyProPoglyAspSerArgLeuProAla..V 61
||||| ||| ||| ||||| ::|||:::|
244 TGTGACCATCTCGGTCTCAGGGGACCTGGAGGCCAGAAATACCTGCTCG 293
61 aGcInlu.....TTPGlyAlaGlnIuLrProValIHISLeu 72
||||| ||| ::|||::|||:::
294 TTTAAACAGGAAGCCCAAGACCTGGGACACACAGAACCCACG..... 337
73 AspSerProAlaIleIySHISGInPheLeuThrGlyAspThrGln... 88
||||| ||| ::|||::|||:::
338 ...GAGCCCAAGACAGAGCCAGATTCTCCATCCATCCATGACAGAGA 384
89 .....GIATGTyTArGcYsArGSeRcIyLeuSerThrGlyTP**G 103
||||| ||| ::|||::|||:::
385 CCATGACAGGAGATACCGCTGTACTACTACAGCCCTGCAGGCTGTGACG 434
103 InLeuSerIyLeuGluLeu 110
:: ||| |||||
435 AGCCGACGACCCCTGACGCTG 457
```

```
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV44826
seq_documentation_block:
ID AAV44826 standard; DNA; 1991 BP.
XX
AC AAV44826;
XX
DT 20-OCT-1998 (first entry)
XX
DE FcR-III coding sequence.
XX
XX Fc receptor-like protein; phagocytosis inducer; rheumatoid arthritis;
KW immune complex related disease; systemic lupus erythematosus; allergy;
KW haemolytic anaemia; thrombocytopenia; anaphylaxis; cancer; lymphoma;
KW leukaemia; infection; immunomodulator; viral entry inhibitor; therapy;
KW FcR-III; ss.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT CDS 73..1944
FT /tag.. a
FT /product.. FcR-III
FT sig_peptide 73..120
FT /tag.. b
FT mat_peptide 121..1941
FT /tag.. c
XX
PN WO9831806-A2.
XX
PD 23-JUL-1998.
XX
PP 20-JAN-1998; 98WO-US01184.
XX
PR 18-JUN-1997; 97US-0049872.
PR 21-JAN-1997; 97US-0034205.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Gentz RL, Murphy M, Ni J, Olsen HS, Ruben SM;
XX
DR WPI: 1998-414105/35.
DR P-PSDB: AAM69233.
XX
XX Nucleic acid encoding Fc receptor-like polypeptides or their
PT fragments - and related vectors, transformed cells and antibodies,
PT useful for treating and diagnosing diseases of the hematopoietic
PT and immune systems
XX
XX Claim 2: Fig 3A; 141pp; English.
XX
XX This sequence encodes the Fc receptor-like III protein (FcR-III) of the
CC invention. Cells containing the DNA are used to express the recombinant
CC protein, and to screen for specific (ant)agonists. The proteins are used
CC to induce phagocytosis, and their (ant)agonists are used to treat immune
CC complex related diseases (e.g. rheumatoid arthritis, systemic lupus
CC erythematosus, haemolytic anaemia, thrombocytopenia, anaphylaxis,
CC allergy, colorectal or breast cancer, lymphoma, leukaemia, infection by
CC intracellular pathogens etc). The antagonists are also useful as
CC immunomodulators and inhibitors of viral (e.g. human immune deficiency or
CC dengue viruses) entry into cells. The proteins may also be used to screen
CC for specific binding agents, i.e. (ant)agonists, for raising antibodies
CC (Ab), and for identification of particular cells or tissues. The Ab can
CC be used therapeutically as antagonists; as assay reagents for diagnostic
CC determination of the levels of expression of the proteins and for
CC affinity purification of the proteins. The DNA and its fragments are
CC useful as hybridisation probes or primers for isolating related genes, in
CC situ hybridisation (chromosome mapping) and diagnostically to measure
CC mRNA expression.
XX
XX Sequence 1991 BP; 430 A; 660 C; 548 G; 353 T; 0 other;
```

alignment_scores:

Quality: 90.50 Length: 138
Ratio: 1.560 Gaps: 6
Percent Similarity: 42.029 Percent Identity: 28.986

alignment_block:

US-09-471-276-831 x AAV44826 ..

Align seg 1/1 to: AAV44826 from: 1 to: 1991

```

1 MetSerMetLeuValAlaPheLeuLeuTrpGlyValThrTrpGlyPro 17
  |||::: ::::: ||||| |||::: |||||
73 ATGACGCCGCCCTCACAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 122
  |||::: ::::: ||||| |||::: |||||
17 oValThrGluAlaAlaIlePheTyrluThrGln***SerLeuTrpAlaG 34
  |||::: ::::: ||||| |||::: |||||
123 CAGGACCCCGCTGCAGGCGGCCCTTCCCAAAACCACCCCTCTGCGCTG 172
  |||::: ::::: ||||| |||::: |||||
34 luserGluHis**LeuTyThrLeuGlyGlnCysAspAlaAspValPro 50
  |||::: ::::: ||||| |||::: |||||
173 AG..... 174
  |||::: ::::: ||||| |||::: |||||
51 GlyProGlyAspSerArgLeuProAlaValGlnGluTrpGly..... 65
  |||::: ::::: ||||| |||::: |||||
175 .....CCAGGC.....TCTGTATCATCAGCTGGGGAGGCC 203
  |||::: ::::: ||||| |||::: |||||
66 .....AlaGlnGlu..... 68
  |||::: ::::: ||||| |||::: |||||
204 CGTGACCATCTGTGTCTCAGGGAGCCCTGGAGGCCAGAGTACCAACTGG 253
  |||::: ::::: ||||| |||::: |||||
69 .....ProValHisLeuAsp.....SerPro 75
  |||::: ::::: ||||| |||::: |||||
254 ATTAAGAGGAGCCCAAGACCCCTTGGACAGAAATACCCACTGGAACTCC 303
  |||::: ::::: ||||| |||::: |||||
76 AlaIleLysHisGlnPheLeuLeuThrGlyAspThrGln.....G1 89
  |||::: ::::: ||||| |||::: |||||
304 AAGACAGAGCCAGATTCATCCATCCATCATGACACAGACCATGGAGG 353
  |||::: ::::: ||||| |||::: |||||
89 yArgTyrrArgCysArgSerGlyLeuSerThrGlyTrp**GlnLeuSerL 106
  |||::: ::::: ||||| |||::: |||||
354 GAGATACCCGCTGCACACTATTTACAGCTCTGCAGGCTGTGCAGAGCC 403
  |||::: ::::: ||||| |||::: |||||
106 ysrLeuLeuGluLeu 110
  |||::: ::::: ||||| |||::: |||||
404 ACCCCCTGGAGCTG 417
  |||::: ::::: ||||| |||::: |||||

```

seq_name: /SID2/gcdata/geneseq/geneseqn/NA1998.DAT:AAV69338

seq_documentation_block:

ID AAV69338 standard: DNA: 2194 BP.

```

XX AC AAV69338;
XX DT 09-FEB-1999 (first entry)
XX DE Human LIR-pbm17 CDNA.
XX KW LIR-pbm17; immunoregulator; leukocyte immunoglobulin-like receptor;
XX KW therapeutic treatment; disorder; ITIM; MHC class I receptor; inhibitor;
XX KW cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
XX KW negative signalling; autoimmune disease; suppressor; LIR; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 67..1962 /*tag_a
XX FT FT /*tag_a
XX FT FT /product= "LIR-pbm17"
XX PM W09848017-A1.
XX PD 29-OCT-1998.

```

```

XX XX 23-APR-1998; 98WO-0508244.
XX PF (LIR) polypeptide LIR-pbm17. This sequence can be administered
XX PR 24-APR-1997; 97US-0842248. therapeutically to treat disorders associated with insufficient/defective
XX PA (IMV ) IMMUNEX CORP. amounts of LIR polypeptide. LIR-P3G2 and certain other LIR family members
XX PI Cosman DJ; contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
XX PS Example 9; Page 87-89; 112pp; English. (ITIMs), whilst other LIR family members lack ITIMs. By analogy with the
XX CC This sequence encodes a novel leukocyte immunoglobulin-like receptor structure and function of known MHC Class I receptor molecules, LIRs
XX CC (LIR) polypeptide LIR-pbm17. This sequence can be administered having ITIMs are inhibitory receptors mediating negative signalling,
XX CC therapeutically to treat disorders associated with insufficient/defective whilst those lacking ITIMs are activatory receptors. Failure of a
XX CC amounts of LIR polypeptide. LIR-P3G2 and certain other LIR family members receptor that mediates negative signalling could result in autoimmune
XX CC contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs diseases, whilst failure of a receptor mediating activatory signalling
XX CC (ITIMs), whilst other LIR family members lack ITIMs. By analogy with the could result in suppressed immune function. They are also useful to
XX CC structure and function of known MHC Class I receptor molecules, LIRs produce probes for detecting LIR nucleic acids or isolating LIR DNA from
XX CC having ITIMs are inhibitory receptors mediating negative signalling, other species.
XX CC whilst those lacking ITIMs are activatory receptors. Failure of a
XX CC receptor that mediates negative signalling could result in autoimmune
XX CC diseases, whilst failure of a receptor mediating activatory signalling
XX CC could result in suppressed immune function. They are also useful to
XX CC produce probes for detecting LIR nucleic acids or isolating LIR DNA from
XX CC other species.
SQ Sequence 2194 BP; 479 A; 729 C; 597 G; 389 T; 0 other:

```

alignment_scores:

Quality: 90.50 Length: 138
Ratio: 1.560 Gaps: 6
Percent Similarity: 42.029 Percent Identity: 28.986

alignment_block:

US-09-471-276-831 x AAV69338 ..

Align seg 1/1 to: AAV69338 from: 1 to: 2194

```

1 MetSerMetLeuValAlaPheLeuLeuTrpGlyValThrTrpGlyPro 17
  |||::: ::::: ||||| |||::: |||||
67 ATGACGCCGCCCTCACAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 116
  |||::: ::::: ||||| |||::: |||||
17 oValThrGluAlaAlaIlePheTyrluThrGln***SerLeuTrpAlaG 34
  |||::: ::::: ||||| |||::: |||||
117 CAGGACCCCGCTGCAGGCGGCCCTTCCCAAAACCACCCCTCTGCGCTG 166
  |||::: ::::: ||||| |||::: |||||
34 luserGluHis**LeuTyThrLeuGlyGlnCysAspAlaAspValPro 50
  |||::: ::::: ||||| |||::: |||||
167 AG..... 168
  |||::: ::::: ||||| |||::: |||||
51 GlyProGlyAspSerArgLeuProAlaValGlnGluTrpGly..... 65
  |||::: ::::: ||||| |||::: |||||
169 .....CCAGGC.....TCTGTATCATCAGCTGGGGAGGCC 197
  |||::: ::::: ||||| |||::: |||||
66 .....AlaGlnGlu..... 68
  |||::: ::::: ||||| |||::: |||||
198 CGTGACCATCTGTGTCTCAGGGAGCCCTGGAGGCCAGAGTACCAACTGG 247
  |||::: ::::: ||||| |||::: |||||
69 .....ProValHisLeuAsp.....SerPro 75
  |||::: ::::: ||||| |||::: |||||
248 ATTAAGAGGAGCCCAAGACCCCTTGGACAGAAATACCCACTGGAACTCC 297
  |||::: ::::: ||||| |||::: |||||
76 AlaIleLysHisGlnPheLeuLeuThrGlyAspThrGln.....G1 89
  |||::: ::::: ||||| |||::: |||||
298 AAGACAGAGCCAGATTCATCCATCCATCATGACACAGACCATGCGAGG 347
  |||::: ::::: ||||| |||::: |||||

```

```
89 yArGTyArGcYsArGSeRcYlEuSeRtHrGlyTrp***GlnLeuSerL 106
|||||
348 GAGATACCCCTGCCACTATTACAGCTCTGCAGCGCTGTCAGACCCAGCC 397
106 yslEuLeuGlnLeu 110
|||||
398 ACCCCCTGCAGCTG 411

seq_name: /SID52/gcdata/geneseq/geneseq/NA2000.DAT:AAA54601

seq_documentation_block:
ID   AAA54601 standard; cDNA; 2194 BP.
XX
AC   AAA54601;
XX
DT   11-APR-2001 (first entry)
XX
DE   Leukocyte immunoglobulin like receptor coding sequence (pbm17).
XX
KW   Leukocyte immunoglobulin like receptor; LIR; gene therapy;
XX   autoimmunity; autoimmune disorders; immune system; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key      Location/Qualifiers
FT   CDS      67..1962
FT           /*tag..a
FT           /product.. Leukocyte immunoglobulin like receptor
XX
PN   WO200068363-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13228.
XX
PR   12-MAY-1999; 99US-0310463.
XX
PA   (IMV ) IMMUNEX CORP.
XX
PI   Cosman DJ, Anderson DM, Borges L;
XX
DR   WPI; 2000-687645/67.
DR   P-PSDB; AAB04177.
XX
PT   Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
PT   treating autoimmune diseases and disease states with suppressed immune
PT   function
XX
PS   Example 9: Page 92-95; 117pp; English.
XX
CC   Host cells transformed with a vector which are capable of
CC   expressing a leukocyte immunoglobulin like receptor (LIR), can be
CC   used to produce the LIR polypeptide. LIR coding sequences may be
CC   used in the gene therapy of disorders mediated directly or
CC   indirectly by defective or insufficient amounts of any of the LIR
CC   polypeptides. The LIR polypeptides can be used to treat autoimmune
CC   diseases and disease states with suppressed immune function.
XX
SQ   Sequence 2194 BP; 479 A; 729 C; 597 G; 389 T; 0 other;

alignment_scores:
Quality: 90.50      Length: 138
Ratio: 1.560      Gaps: 6
Percent Similarity: 42.029      Percent Identity: 28.986

alignment_block:
US-09-471-276-831 x AAA54601 ..

Align seg 1/1 to: AAA54601 from: 1 to: 2194
1 MetSerMetLeuValValPheLeuLeuTTrpGlyValTTrpGlyPr 17
```

```
|||||
67 ATGACCCCGCCCTGCACAGCCCTGCTCCTGGCTGAGTCTGGGCC 116
17 oValTTrpGlnAlaIAlaIlePheTyrGlnThrGln***SerLeuTrpAlaG 34
|||||
117 CAGACCCCGCGTGCAGGCAAGGCCCTTCCCAAAACCACCTCTGGGCTG 166
34 luserGlnHis**LeuYsTTrLeuGlyGlnCysAspAlaAspValPro 50
||
167 AG.....
51 GlyProProGlyAspSerArgLeuProAlaValGlnGlnTTrpGly..... 65
|||||
169 .....CCAGGC.....TCTGTATCATCAGCTGGGGAGGCC 197
66 .....AlaGlnGlu..... 68
198 CGTGACCATCTGCTGCAGGGAGCCTGAGGCCAGAGATACCAACTG 247
69 .....ProValHisLeuAsp.....SerPro 75
248 ATTAAGAGCGAAGCCCAAGACCCCTTGACAGAAATACCCACTGGAACCC 297
76 AlaIleYsHisGlnPheLeuLeuTTrpGlyAspTTrGln.....G1 89
298 AGAACAAGAGCCAGATTCATCCATCCATGACAGACACCATGCAG 347
89 yArGTyArGcYsArGSeRcYlEuSeRtHrGlyTrp***GlnLeuSerL 106
|||||
348 GAGATACCCCTGCCACTATTACAGCTCTGCAGCGCTGTCAGACCCAGCC 397
106 yslEuLeuGlnLeu 110
|||||
398 ACCCCCTGCAGCTG 411

seq_name: /SID52/gcdata/geneseq/geneseq/NA1998.DAT:AAV69334

seq_documentation_block:
ID   AAV69334 standard; DNA; 2446 BP.
XX
AC   AAV69334;
XX
DT   09-FEB-1999 (first entry)
XX
DE   Human LIR-pbm36-2 cDNA.
XX
KW   LIR-pbm36-2; immunoregulator; leukocyte immunoglobulin-like receptor;
KW   therapeutic; treatment; disorder; ITIM; MHC class I receptor; inhibitor;
KW   cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
KW   negative signalling; autoimmune disease; suppressor; LIR; ss.
XX
OS   Homo sapiens.
XX
FH   Key      Location/Qualifiers
FT   CDS      171..1040
FT           /*tag..a
FT           /product.. "LIR-pbm36-2"
XX
PN   WO9848017-A1.
XX
PD   29-OCT-1998.
XX
PF   23-APR-1998; 98WO-US08244.
XX
PR   24-APR-1997; 97US-0842248.
XX
PA   (IMV ) IMMUNEX CORP.
XX
PI   Cosman DJ;
XX
DR   WPI; 1998-609990/51.
DR   P-PSDB; AAW82548.
XX
```

PT Leukocyte immunoglobulin-like receptor, LIR, polypeptides - useful
PT e.g. for treating autoimmune diseases or disease states associated
PT with suppressed immune function

PS Example 9; Page 71-73; 112pp; English.

CC This sequence encodes a novel leukocyte immunoglobulin-like receptor
CC (LIR) polypeptide LIR-ph36-2. This sequence can be administered
CC therapeutically to treat disorders associated with insufficient/defective
CC amounts of LIR polypeptide. LIR-ph36-2 and certain other LIR family members
CC contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
CC (ITIMs), whilst other LIR family members lack ITIMs. By analogy with the
CC structure and function of known MHC Class I receptor molecules, LIRs
CC having ITIMs are inhibitory receptors mediating negative signalling,
CC whilst those lacking ITIMs are activatory receptors. Failure of a
CC receptor that mediates negative signalling could result in autoimmune
CC diseases, whilst failure of a receptor mediating activatory signalling
CC could result in suppressed immune function. They are also useful to
CC produce probes for detecting LIR nucleic acids or isolating LIR DNA from
CC other species.

Sequence 2446 BP; 611 A; 692 C; 557 G; 586 T; 0 other;

[illegible]

```
alignment_block:
US-09-471-276-831 x AAV69334 ..
```

Align seg 1/1 to: AAV69334 from: 1 to: 2446

```

1 MeSerMetLeuVal1ValPheLeuLeuLeuTrpGluValThrTrpGlyTr 17
   |||::: ::::: |||::: |||::: |||:::
171 ATACACCCCATCCTCAGAGCTCTGATCTGTCTCGGGCTAGCTGGGCC 220
   |||::: ::::: |||::: |||::: |||:::
17 ovalThrGluAla1AlaIlePheTrpGluThrGln**SerLeuTrpAlaG 34
   | |||::: ::::: |||::: |||::: |||:::
221 CCGGACCCAGCATGCGAGGACGAGGACCTCCCAAGCCACACATCTGGGCTG 270
   |||::: ::::: |||::: |||::: |||:::
34 IuSerGluHis**LeuLysThrLeuGly..GlnLysAspAlaAspValP 50
   || |||::: ||| ||| ||| ||| ||| |||
271 AGCGAGGCTCTGTGATC...ACCCAGGGAGATCCCGTGACCCCTGTGGTGT 317
   || |||::: ||| ||| ||| ||| ||| |||
50 roGluProProlGlyAspSerArgLeuProAlaVal..GlnLys..... 63
   ||| ||| ||| ||| ||| ||| ||| ||| |||
318 CAGGGGATCTCTGGAGACCCAGGAGTACCGTCTGTATAGAGAAAGAAAC 367
   ||| ||| ||| ||| ||| ||| ||| ||| |||
64 .....TrpGluYalGlnGluProValHisLeuAspSerProAlaIleL 78
   ||| ||| ||| ||| ||| ||| ||| ||| |||
368 AGCACCCCTGGATTCACAGGATCCACAGAGATY.....GTACAGA 408
   ||| ||| ||| ||| ||| ||| ||| ||| |||
78 yshSgIuPhe.....LeuLeuTrpGluYaspThrGlnGlyArgTyr 91
   || ||| ||| ||| ||| ||| ||| ||| ||| |||
409 AGGGCCAGTTCCCATCCTCCGTCATCATCTGGGAACACACCGGGCGGTAT 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
92 ArgCysArgSerGlyLeuSerThr...GlyTrp**GlnLeuSerLysLe 107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
459 CGGTGTTTACAGCTAGCACACTGCACAGCTGCTCAGACCCAGGTACCC 508
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 uLeuGlnLeu 110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
509 CCTGGAGCTG 518

```

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2000.DAT:AAA54597

```
seq_documentation_block:
ID   AAA54597 standard; cDNA; 2446 BP.
XX
AC   AAA54597;
XX
```

DT 11-APR-2001 (first entry)

DE Leukocyte immunoglobulin like receptor coding sequence (pbm36-2).

KW Leukocyte immunoglobulin like receptor; LIR; gene therapy;

KW autoimmunity; autoimmune disorders; immune system; human; ds.
...

05 Homo sapiens.

FH	Key	Location/Qualifiers
171	1010	

FT	CDS	171..1040
FT		4433

```

E1 / *tag= a
E2 /product= Immunoglobulin like receptor

```

XXIX

PN WO200068383 -

XX

PD 16-NOV-2000.

[illegible]

PF 12-MAY-2000; 2000WO-US132228.
XX

XX 12-MAY-1999. 09115-
PR

PR 12-MAY-1999; 95
XX

AA
PA (IMMV) IMMUNEX CORP.

XX 1

PI Cosman, D

XX
XX

DR WPI; 2000-687645/67.
DCC B-SCDD: AAD0A173

DR P-PSDB; AAB04173.
XY

	Taukocyte immunoglobulin-like receptor (TIR) nucleic acids - useful
XX	
PT	

PT Creating autoimmune diseases and disease states with suppressed

PT function

XXXX

PS Example 9; Page 77-79; 117pp; English.

XX
XX
"-----"

CC Host cells transformed with vector which are capable of
CC expressing a leukocyte immunoglobulin like receptor (LIR), can be
CC used to produce the LIR polypeptide. LIR coding sequences may be
CC used in the gene therapy of disorders mediated directly or
CC indirectly by defective or insufficient amounts of any of the LIR
CC polypeptides. The LIR polypeptides can be used to treat autoimmune
CC diseases and disease states with suppressed immune function.
XX
XX Sequence 2446 BP; 611 A; 692 C; 557 G; 586 T; 0 other;

alignment_scores:

Quality:	89.50	Length:	121
Ratio:	1.377	Gaps:	7

```

alignment-block:
US-09-471-276-831 x AAA54597 ..

Align seg 1/1   to: AAA54597   from: 1   to: 2446

1 MetSerMetIeuValValPheLeuLeuTrpGlyValThrTrpGlyPr 17
|||::: :::::::::: |||::: |||::: |||:::
171 ATGACCCCCATCTCCACGGCTCCTGATCTGTCTCGGCGTAGTCTGGGCC 220
17 OValThrGlAlaIalIlePheTyGIuThrGln**SerLeuTPAlac 34
| | | | | : : : : : : : : : : : : : : : : : : : : : :
221 CCGAGCCCATCTGCAGCGAGGACCCTCCCAAGCCACACTTCCTGGGCTG 270
34 IuSerGIuHis**LeuTyStrIeuGly.GIncysAspAlaAspValp 50
|| | | | | | | | | | | | | | | | | | | | | | |
271 AGCCAGGCTCTGTGATC...ACCAGGGAGATCCCGTGACCTCTGGGTGT 317
50 rOcGIProPIroGIVASerArqLeuProAlaVal.GIInGU..... 63
||| | | | | | | | | | | | | | | | | | | | | |
318 CAGGGGATCTCGAGAACCCAGAGATACCTCTGTATAGAGAAAAGAAMAC 367
64 .....TPrlYalagIngIUProValHisLeuAspSerProAlaIlel 78

```



```

PM WO200100810-A1.
XX
XX 04-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US18152.
XX
XX 30-JUN-1999; 99US-0345468.
XX PR 06-DEC-1999; 99US-0454824.
XX PR 14-FEB-2000; 2000US-0503387.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Busfield SJ, Villalaj J, Jandrot-Perrus M, Vainchenker W, Gili DS;
XX Qian MD, Kingsbury G;
XX WPI: 2001-080877/09.
XX P-PSDB; AAB61263.
XX
XX New genes encoding human platelet-expressed collagen receptor,
XX glycoprotein VI, and its modulators, useful for preventing, treating
XX and diagnosing hemorrhagic disorders, thrombotic diseases and
XX immunological disorders -
XX
XX Disclosure: Fig 3A-3D; 227pp; English.
XX
XX The present sequence is given in a specification relating to an isolated
XX nucleic acid molecule encoding a platelet membrane glycoprotein receptor
XX glycoprotein VI (GPVI), also called TANGO 268. The GPVI polynucleotides
XX and polypeptides and their modulators, e.g. antisense nucleic acids,
XX ribozymes and antibodies, are useful for preventing, treating and
XX diagnosing disorders associated with aberrant expression or activity of
XX GPVI. These disorders include bleeding disorders
XX (e.g. thrombocytopaenia), blood vessel injury, thrombotic disorders
XX (e.g. thrombotic occlusion of the coronary arteries), haemorrhagic
XX disorders, coronary artery and cerebral artery diseases (e.g. stroke and
XX ischaemia), cardiovascular diseases (e.g. atherosclerosis and myocardial
XX infarction), immunological diseases (e.g. platelet disorder) and
XX embryonic liver disorders. Preferably they are used to prevent acute
XX cardiac ischaemia following angioplasty and metastatic cancers,
XX especially of the colon and liver.
XX
XX Sequence 2170 BP; 478 A; 716 C; 591 G; 385 T; 0 other;
XX
XX
XX alignment_scores:
XX      Quality: 87.50      Length: 138
XX      Ratio: 1.509      Gaps: 6
XX      Percent Similarity: 42.029      Percent Identity: 28.261
XX
XX alignment_block:
XX US-09-471-276-831 x AAF29472 ..
XX
XX Align seg 1/1 to: AAF29472 from: 1 to: 2170
XX
XX 1 MeiserMetLeuValValPheLeuLeuTrpGlyValThrTrpGlyPr 17
XX ||||| :|||: ||||| |||||: |||||
XX 50 ATGACGCCGCCCTCACAGCCCTCTGCTGCTGGCTGAGTGGGCC 99
XX
XX 17 ovaTrhGluAlaAlaIlePheTrpGluTrpGln**SerLeuTrpAlaG 34
XX | :|||: ||||| :|||: |||||
XX 100 CAGACCCGGGTGACGACGAGGCCCTTCCCAACACACCTCTGGGCTG 149
XX
XX 34 lueSerGluHis**LeuLysThrLeuGlyGlnCysAspAlaAspValPro 50
XX ||
XX 150 AG.....CCAGGC.....TCTGTGATCAGCTGGGGGAGCCCC 151
XX
XX 51 GlyProGlyAspSerArgLeuProAlaValGlnGluTrpGly..... 65
XX ||||| :|||: ||||| :|||: |||||
XX 152 .....CCAGGC.....TCTGTGATCAGCTGGGGGAGCCCC 180
XX
XX 66 .....AlaGlnGlu..... 68
XX |||||
XX 181 CGTGACCATCTGCTGTGAGGGAGACCTGAGAGGCCACGAGTACGCGACTGG 230

```

```

69 .....ProValHisLeuAsp.....SerPro 75
XX ||| ||||| :|||
XX 231 ATAAAGAGGAGACCCACAGAGCCCTTGACAGAAATACCCACTGGAACCC 280
XX ||| :||| :||| :|||
XX 76 AlaIleLysHisGlnPheLeuLeuTrpGlyAspThrGln.....G1 89
XX ||| :||| :||| :|||
XX 281 AAGAACAAAGCCAGATCTCCATCCATCCATGACAGACCACTGGGG 330
XX ||||| :||| :||| :|||
XX 89 YArgTYrArgCysArgSerGlyLeuSerThrGlyTrp**GlnLeuSerL 106
XX ||||| :||| :||| :|||
XX 331 GAGATACCCCTGGCCACTTATACAGCTCTGCAGAGCTGTGTCAGAGCCAGCG 380
XX |||||
XX 106 YsLeuLeuGluLeu 110
XX |||||
XX 381 ACCCCCTGGAGCTG 394
XX
XX seq_name: /Sids2/gcgsdata/geneseq/geneseqn/NA1998.DAT:AAV69335
XX
XX seq_documentation_block:
XX ID AAV69335 standard; DNA; 1910 BP.
XX
XX AAV69335;
XX
XX 09-FEB-1999 (first entry)
XX
XX Human LIR-pbm36-4 cDNA.
XX
XX LIR-pbm36-4; immunoregulator; leukocyte immunoglobulin-like receptor;
XX therapeutic; treatment; disorder; ITIM; MHC class I receptor; inhibitor;
XX cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
XX negative signalling; autoimmune disease; suppressor; LIR; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 183..1652
XX FT /*tag= a
XX FT /product= "LIR-pbm36-4"
XX
XX PN WO9848017-A1.
XX
XX 29-OCT-1998.
XX
XX 23-APR-1998; 98WO-US08244.
XX
XX 24-APR-1997; 97US-0842248.
XX
XX (IMMV ) IMMUNEX CORP.
XX
XX Cosman DJ;
XX
XX WPI: 1998-609990/51.
XX
XX P-PSDB; AAW82549.
XX
XX Leukocyte immunoglobulin-like receptor, LIR, polypeptides - useful,
XX e.g. for treating autoimmune diseases or disease states associated
XX with suppressed immune function
XX
XX Example 9: Page 74-77; 112pp; English.
XX
XX This sequence encodes a novel leukocyte immunoglobulin-like receptor
XX (LIR) polypeptide LIR-pbm36-4. This sequence can be administered
XX therapeutically to treat disorders associated with insufficient/defective
XX amounts of LIR polypeptide. LIR-p362 and certain other LIR family members
XX contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
XX (ITIMs), whilst other LIR family members lack ITIMs. By analogy with the
XX structure and function of known MHC Class I receptor molecules, LIRs
XX having ITIMs are inhibitory receptors mediating negative signalling,
XX whilst those lacking ITIMs are activatory receptors. Failure of a
XX receptor that mediates negative signalling could result in autoimmune
XX diseases, whilst failure of a receptor mediating activatory signalling
XX could result in suppressed immune function. They are also useful to

```

CC Produce probes for detecting LIR nucleic acids or isolating LIR DNA from
CC other species.

SO Sequence 1910 BP; 408 A; 588 C; 509 G; 405 T; 0 other:

alignment_scores:

Quality:	84.50	Length:	121
Ratio:	1.320	Gaps:	7
Percent Similarity:	52.893	Percent Identity:	33.884

alignment_block:

US-09-471-276-831 x AAV69335 ..

Align seg 1/1 to: AAV69335 from: 1 to: 1910

```

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
183 ATGACCCCTCTGATTCAGTCTGTCAGGCTGAGTGGGCC 232
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
17 oValThrGluAlaAlaIlePheThrGluThrGln***SerLeuTrpAlaG 34
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
233 CCGGACCCAGTGTACGACGAGCCCTCCCAAGCCACACTGTGGCTG 282
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
34 luserGluHis**LeuLysThrLeuGly..GlnCysAspAlaAspValP 50
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
283 AGCGAGGCTCTGTGATC...ACCCAGGGGAGTCCCGTGGCTGTGTGT 329
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
50 roGlyProProGlyAspSerArgLeuProAlaVal.GlnGlu..... 63
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
330 CAGGGGATCTCTGAGACCCAGCAGTACCGCTGTATAGACAAAGAAAC 379
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
380 AGCACCCTGGATTACAGGGATCCACAGAGATT.....GTGAGA 420
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
78 yshIsGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArgTyr 91
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
421 AGGGCCAGTTCCTCCCATCCATCCATCCATCCGAGACACACAGGGCGGTAT 470
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
92 ArgCysArgSerGlyLeuSerThr...GlyTrp**GlnLeuSerLysIle 107
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
471 CGCTGTTTCTACGCTAGCCACACTGCAGCGCTGTGACAGCCAGTGACCC 520
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
107 uLeuGluLeu 110
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
521 CCTGGAGCTG 530
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::

```

seq_name: /SID52/gcdata/geneseq/geneseqn/NA2000.DAT:AAA54598

seq_documentation_block:

ID AAA54598 standard; CDNA; 1910 BP.

AC AAA54598;

DT 11-APR-2001 (first entry)

DE Leukocyte immunoglobulin like receptor coding sequence (pbm36-4).

KW Leukocyte immunoglobulin like receptor; LIR; gene therapy;
KW autoimmunity; autoimmune disorders; immune system; human; ds.

OS Homo sapiens.

FI key Location/Qualifiers

FT CDS 183..1652

FT /*tag" a

FT /product" Leukocyte immunoglobulin like receptor

XX WO200068383-A2.

XX 16-NOV-2000.

XX

PF 12-MAY-2000; 2000WO-US13228.

PR 12-MAY-1999; 99US-0310463.

PA (IMMUNEX) IMMUNEX CORP.

PI Cosman DJ, Anderson DM, Borges L;

DR WPT: 2000-687645/67.

DR P-PSDB: AAB04174.

PT Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for
PT treating autoimmune diseases and disease states with suppressed immune
PT function

PS Example 9; Page 80-83; 117pp; English.

CC Host cells transformed with a vector which are capable of
CC expressing a leukocyte immunoglobulin like receptor (LIR), can be
CC used to produce the LIR polypeptide. LIR coding sequences may be
CC used in the gene therapy of disorders mediated directly or
CC indirectly by defective or insufficient amounts of any of the LIR
CC polypeptides. The LIR polypeptides can be used to treat autoimmune
CC diseases and disease states with suppressed immune function.

SO Sequence 1910 BP; 408 A; 588 C; 509 G; 405 T; 0 other:

alignment_scores:

Quality:	84.50	Length:	121
Ratio:	1.320	Gaps:	7
Percent Similarity:	52.893	Percent Identity:	33.884

alignment_block:

US-09-471-276-831 x AAA54598 ..

Align seg 1/1 to: AAA54598 from: 1 to: 1910

```

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
183 ATGACCCCTCTGATTCAGTCTGTCAGGCTGAGTGGGCC 232
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
17 oValThrGluAlaAlaIlePheThrGluThrGln***SerLeuTrpAlaG 34
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
233 CCGGACCCAGTGTACGACGAGCCCTCCCAAGCCACACTGTGGCTG 282
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
34 luserGluHis**LeuLysThrLeuGly..GlnCysAspAlaAspValP 50
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
283 AGCGAGGCTCTGTGATC...ACCCAGGGGAGTCCCGTGGCTGTGTGT 329
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
50 roGlyProProGlyAspSerArgLeuProAlaVal.GlnGlu..... 63
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
330 CAGGGGATCTCTGAGACCCAGGAGTACCGCTGTATAGACAAAGAAAC 379
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
380 AGCACCCTGGATTACAGGGATCCACAGAGATT.....GTGAGA 420
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
78 yshIsGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArgTyr 91
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
421 AGGGCCAGTTCCTCCCATCCATCCATCCATCCGAGACACACAGGGCGGTAT 470
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
92 ArgCysArgSerGlyLeuSerThr...GlyTrp**GlnLeuSerLysIle 107
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
471 CGCTGTTTCTACGCTAGCCACACTGCAGCGCTGTGACAGCCAGTGACCC 520
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
107 uLeuGluLeu 110
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::
521 CCTGGAGCTG 530
  |||::: ::|||::: |||::: ::|||::: |||::: |||:::

```

seq_name: /SID52/gcdata/geneseq/geneseqn/NA1998.DAT:AAV69339

seq_documentation_block:

XX	AAV69339 standard; DNA: 2061 BP.
XX	AAV69339;
XX	
XX	09-FEB-1999 (first entry)
XX	
DE	Human LIR-pbmnew cDNA.
XX	
KW	LIR-pbmnew; immunoregulator; leukocyte immunoglobulin-like receptor;
KW	therapeutic; treatment; disorder; ITIM; MHC class I receptor; inhibitor;
KW	cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
KW	negative signalling; autoimmune disease; suppressor; LIR; ss.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Location/Qualifiers
FT	67..1839
FT	/*tag= a
FT	/product= "LIR-pbmnew"
XX	
PN	MO9848017-A1.
XX	
PD	29-OCT-1998.
XX	
PE	23-APR-1998; 98WO-US08244.
XX	
PR	24-APR-1997; 97US-0842248.
XX	
PA	(IMMV) IMMUNEX CORP.
P1	
P1	Cosman DJ;
XX	
DR	WPI; 1998-609990/51.
DR	P-PSDB; AAM82553.
XX	
P1	Leukocyte immunoglobulin-like receptor, LIR, polypeptides - useful,
P1	e.g. for treating autoimmune diseases or disease states associated
PT	with suppressed immune function
XX	
PS	Example 9; Page 92-94; 112pp; English.
XX	
CC	This sequence encodes a novel leukocyte immunoglobulin-like receptor
CC	(LIR) polypeptide LIR-pbmnew. This sequence can be administered
CC	therapeutically to treat disorders associated with insufficient/defective
CC	amounts of LIR polypeptide. LIR-P3g2 and certain other LIR family members
CC	contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
CC	(ITIMs), whilst other LIR family members lack ITIMs. By analogy with the
CC	structure and function of known MHC Class I receptor molecules, LIRs
CC	having ITIMs are inhibitory receptors mediating negative signalling,
CC	whilst those lacking ITIMs are activatory receptors. Failure of a
CC	receptor that mediates negative signalling could result in autoimmune
CC	diseases, whilst failure of a receptor mediating activatory signalling
CC	could result in suppressed immune function. They are also useful to
CC	produce probes for detecting LIR nucleic acids or isolating LIR DNA from
CC	other species.
XX	
SO	Sequence 2061 BP; 431 A; 684 C; 558 G; 388 T; 0 other;

```

alignment_scores:
    Quality:      84.50      Length:      139
    Ratio:        1.207      Gaps:      9
    Percent Similarity:  50.360      Percent Identity:  30.935

alignment_block:
US-09-471-276-831 x AAV69339      ..      .

Align seg 1/1 to: AAV69339 from: 1 to: 2061

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
|||||:      :|||:|||||:|||||:|||||
67 ATGACCCCTCACCCCTTCAGTCCGATTTGCCCTCGGCGCTAGTGTGGCCCC 118

```

[illegible]

seq_name: /S10524/gcgcdata/geneseq/geneseqn/NA2000.DAT:AAA54602	
seq_documentation_block:	
ID AAA54602 standard: cDNA; 2061 BP.	
AC AAA54602;	
XX	
DT * 11-APR-2001 (first entry)	
XX	
DE Leukocyte immunoglobulin like receptor coding sequence (pbmmew).	
XX	
KW Leukocyte immunoglobulin like receptor; LIR; gene therapy;	
KW autoimmunity; autoimmune disorders; immune system; human; ds.	
XX	
OS Homo sapiens.	
XX	
EH Key Location/Qualifiers	
FT CDS 67..1839	
FT /tag_a	
FT /product= Leukocyte immunoglobulin like receptor	
XX	
PN W0200068383-A2.	
XX	
PD 16-NOV-2000.	
XX	
PE 12-MAY-2000; 2000MO-US13228.	
XX	
PR 12-MAY-1999; 99US-0310463.	
XX	
PA (IMMUNEX) IMMUNEX CORP.	
XX	
PI Cosman DJ, Anderson DM, Borges LJ;	
XX	
DR WPI: 2000-687645/67.	
DR P-PSDB: AAB04178.	
XX	
PT Leukocyte immunoglobulin-like receptor (LIR) nucleic acids, useful for	
PT treating autoimmune diseases and disease states with suppressed immune	
PT function	
XX	
PS Disclosure; Page 97-100; 117pp; English.	


```

KW Human: gp49; HM18; HM43; Immunoglobulin; immune response; mast cell;
KW bone marrow; cell-surface member; FcERI; ss.
XX
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT CDS 63..1382
FT /tag "a
FT /product "gp49 HM43"
XX
XX MO9809638-A1.
XX
XX 12-MAR-1998.
XX
XX 05-SEP-1997; 97WO-US15586.
XX
XX 06-SEP-1996; 96US-0025846.
XX
XX (BCHM ) BRIGHAM & WOMENS HOSPITAL.
XX
XX Arm JP, Austen KF, Castells MC, Katz HR;
XX
XX WPT: 1998-193318/17.
XX
XX P-PSDB: AAW53464.
XX
XX Cell-surface member of immunoglobulin super-family, human gp49 -
XX useful to treat undesired immune responses, especially mast
XX cell-related diseases
XX
XX Clatm 5; Flg 2A-B; 62pp; English.
XX
XX The present sequence encodes human gp49 polypeptide HM43. The present
XX invention also describes: (1) a fusion polypeptide comprising a human
XX gp49 component and FcERI or a detectable marker; (2) a recombinant
XX nucleic acid encoding human gp49 or a human gp49-FcERI fusion as above;
XX (3) a cell or vector comprising the recombinant nucleic acid as in (2),
XX and (4) an antibody which selectively binds to gp49. Mammalian gp49 or
XX its related DNA can be used to treat an undesired immune response,
XX especially a mast cell-related disease.
XX
XX Sequence 1593 BP; 354 A; 485 C; 425 G; 329 T; 0 other:
SQ

```

alignment_scores:

Quality:	81.50	Length:	121
Ratio:	1.254	Gaps:	7
Percent Similarity:	53.719	Percent Identity:	32.231

alignment_block:

```

US-09-471-276-831 x AAV23274 ..

```

Align seg 1/1 to: AAV23274 from: 1 to: 1593

```

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
   |||: : : : : : : : : : : : : : : : : : : : : : : :
63 ATGACCTCCATCTCCACGGCTGATGTCTCGGCTGACCTGACCC 112
   | : : : : : : : : : : : : : : : : : : : : : : :
17 ovaTthGluAlaIlePheTyrGluThrGln**SerLeuTrpAlaG 34
   | : : : : : : : : : : : : : : : : : : : : : : :
113 CAGACCCACGCTGACGAGCGCCCTCCCAAGCCACCCCTCGGCTG 162
   | : : : : : : : : : : : : : : : : : : : : : : :
34 LuseRGIuHIS**IeulYsThrLeuGly..GInCYsAspAlaAspAlP 50
   || : : : : : : : : : : : : : : : : : : : : : : :
163 AGCCAGGCTCTGTGATC...ACCCAGGAGACTCTGTGACCCCTGAG 209
   || : : : : : : : : : : : : : : : : : : : : : : :
50 roGlyProPogIyAspSerArgLeuProAlaVal.GInGlu..... 63
   || : : : : : : : : : : : : : : : : : : : : : : :
210 CAGGAGGAGCTGAGAGCGACGACTACATCTATATGACAAAGAAAC 259
   || : : : : : : : : : : : : : : : : : : : : : : :
64 .....TrpGlyAlaGInGluProValHISLeuAspSerProAlaIle 78
   || : : : : : : : : : : : : : : : : : : : : : : :
260 AGCACTCTGATTACACGATCCACAGAGAGCTT.....CTGACAGA 300

```

```

78 yuHISGInPhe.....LeuLeuThrGlyAspThrGInGlyArgTyr 91
   || : : : : : : : : : : : : : : : : : : : : : : :
301 AGGGCCAGTTCCTATCCATCCTATCCATCCTGGGAGACATCGAGCGGTAT 350
   || : : : : : : : : : : : : : : : : : : : : : : :
92 ArgCYsArgSerGlyLeuSerThr...GlyTrp***GInLeuSerLysIe 107
   || : : : : : : : : : : : : : : : : : : : : : : :
351 TGCCTGATCTATGCGACGACACTGACGCGCTCTCAGAGAGAGTGACCC 400
   || : : : : : : : : : : : : : : : : : : : : : : :
107 uLeuGluLeu 110
   || : : : : : : : : : : : : : : : : : : : : : : :
401 CCTGAGAGCTG 410

```

seq_name: /SIDS2/gcdata/geneseq/NA1998.DAT:AAV69332

seq_documentation_block:

```

ID AAV69332 standard; DNA; 1605 BP.
XX
XX AAV69332;
XX
XX 09-FEB-1999 (first entry)
XX
XX Human LIR-pbm25 cDNA.
XX
XX LIR-pbm25; immunoregulator; leukocyte immunoglobulin-like receptor; LIR;
XX therapeutic; treatment; disorder; ITIM; MHC class I receptor; inhibitor;
XX cytoplasmic immunoreceptor tyrosine-based inhibitory motif; activator;
XX negative signaling; autoimmune disease; suppressor; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 93..1412
XX /tag "a
XX /product "LIR-pbm25"
XX
XX MO9848017-A1.
XX
XX 29-OCT-1998.
XX
XX 23-APR-1998; 98WO-US08244.
XX
XX 24-APR-1997; 97US-0842248.
XX
XX (IMWV ) IMMUNEX CORP.
XX
XX Cosman DJ;
XX
XX WPT: 1998-609990/51.
XX
XX P-PSDB: AAW82546.
XX
XX Leukocyte immunoglobulin-like receptor, LIR, polypeptides - useful,
XX e.g. for treating autoimmune diseases or disease states associated
XX with suppressed immune function
XX
XX Example 9; Page 62-64; 112pp; English.
XX
XX This sequence encodes a novel leukocyte immunoglobulin-like receptor
XX (LIR) polypeptide LIR-pbm25. This sequence can be administered
XX therapeutically to treat disorders associated with insufficient/defective
XX amounts of LIR polypeptide. LIR-P3G2 and certain other LIR family members
XX contain cytoplasmic immunoreceptor tyrosine-based inhibitory motifs
XX (ITIMs), whilst other LIR family members lack ITIMs. By analogy with the
XX structure and function of known MHC Class I receptor molecules, LIRs
XX having ITIMs are inhibitory receptors mediating negative signalling,
XX whilst those lacking ITIMs are activatory receptors. Failure of a
XX receptor that mediates negative signalling could result in autoimmune
XX diseases, whilst failure of a receptor mediating activatory signalling
XX could result in suppressed immune function. They are also useful to
XX produce probes for detecting LIR nucleic acids or isolating LIR DNA from
XX other species.
XX
XX Sequence 1605 BP; 347 A; 495 C; 430 G; 333 T; 0 other:
SQ

```


XX		Human monocytic gene designated YTB01.
DE		
XX		
KW		Human; type I transmembrane protein; immunoglobulin-like domain;
KM		FDP03; activated monocyte; YE01; KTE03; control; development;
KV		differentiation; mammalian immune system; treatment; cancerous condition
KW		degenerative condition; autoimmune response; transplantation rejection;
KM		graft versus host disease; inflammatory condition; detection; diagnosis;
XX		drug screening; ss.
XX		
OS	Homo sapiens.	
FM		
FT	Key	Location/Qualifiers
CDS		81..1400
FT	/tag "a	
FT	/product YYB01	
PN	MO9824906-AZ.	
PD	11-JUN-1998.	
PP	05-DEC-1997;	97WO-US21101.
PR	21-MAR-1997;	97US--0041279.
PR	06-DEC-1996;	96US--0032252.
PR	09-DEC-1996;	96US--0762187.
PA	16-DEC-1996;	96US--0033181.
SCHE) SCHERING CORP.		
P1	Adema GJ, Gorman DM, Lanier LL, McClanahan TK, Meyard L;	
P1	Phillips JH, Zurawski G, Zurawski SM;	
DR	MP1; 1998-333325/29.	
DR	P-PADB; AAM62777.	
PT	New isolated activated monocyte cell gene(s) - used to develop	
PT	products for treating e.g. cancer, degenerative conditions,	
PT	autoimmune responses, transplant rejection or inflammatory	
PT	conditions	
PS	Disclosure: Pages 71-73; 104pp: English.	
CC		
CC	AAV38992-96 and AAV38608 represent human monocyte genes, collectively	
CC	designated KTE03. The genes are found in activated monocytes. The	
CC	specification also describes other proteins encoded by activated	
CC	monocytes, which are designated PPF03 and YE01. The genes function in	
CC	controlling development, differentiation, and/or physiology of the	
CC	mammalian immune system. The products can be used for treating abnormal	
CC	proliferation, regeneration, degeneration or atrophy. They can be used	
CC	for treating e.g. cancerous conditions, degenerative conditions,	
CC	autoimmune responses, transplantation rejection, graft versus host	
CC	disease, or inflammatory conditions. The products can also be used for	
CC	detection, diagnosis and drug screening.	
SQ	Sequence 1620 BP; 361 A; 497 C; 434 G; 328 T; 0 other:	
alignment_scores:		
Quality:	80.50	Length: 121
Ratio:	1.238	Gaps: 7
Percent Similarity:	53.719	Percent Identity: 32.231
alignment_block:		
US-09-471-276-831 x AAV38992 ..		
Align seg 1/1 to: AAV38992 from: 1 to: 1620		
1 MetSerMetLeuVal[Val]PheLeuLeuTrpGj[ValThrTrpGlyc 17		
: ::: ::: ::: :::		
81 ATACCCCCTCCTCACGCCTCATCTGTCTCGCGTGAAGCCTGACGCC 130		
17 ovalThrgUAlaAalAlePhetYrGluthRgin***SerLeuTPALag 34		

```

| | | | | : : : : : |||||
131 CAGGACCCACGTGCAGAGCGGCCCTCCCAAGCCACCTCTGGGTG 180
34 lusecCluhis**LeulysThrleucly..GlncysaspIaspylP 50
   ||      ::    |||  ||  ::||| |||
181 AGCAGAGCCTCTGTATC...ACCNAAGGAGTCTGTACCCCTCAGTG 227
50 roglyProProclYasPserArgleuProAlaVal.Glnglu..... 63
   ||| ||||| : : : : :
228 CAGGGAGCCTGTGACAGCCGAGGTACCATCTATVTTATACGAAMAAC 277
64 .....TrpglyAlaIngIuProval.HlsleuaspserProAlaIel 78
   ||| : : : : : ||| : : : ||
278 AACACTCTGGATTACACGGATCCCACAGAGCTT.....GTGAAGA 318
78 yslscInPhe.....LeuleuthrGlyaspThrGlnclYArgtyr 91
   || ||||| : : : : : |||||
319 AAGGGCAGTTCCTCCCATCTATCCATCACCTGGGAACATGACAGGGCG 368
92 ArgCyAsArgSerGlyLeuSerThr...GlyTrp**GlnLeuSerLysLe 107
   ||| ||| ||| ||| ||| : : :
369 TCCTGTATCTATGTGGCAGCCACACTGCAGGCCCTCAGAGCAGTGCACC 418
107 uleugluLeu 110
   ||||| |||
419 CCTGGAGCTG 428

seq_name: /SIBS2/gcgdatu/geneseq/geneseqn/NA2001.DAT:AA151413

seq_documentation_block:
ID AA151413 standard; DNA; 311 BP.
AC AA151413;
XX
DF 17-OCT-2001 (first entry)
XX
DE Probe #2009 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
OS Homo sapiens.
XX
PN MO200157272-A2.
PD
XX
09-AUG-2001.
XX
PF 30-JAN-2001; 2001MO-US00653.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632365.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024253.
XX
PA (MOL-E-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
DR WP1; 2001-488897/53.
XX
Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25: SEQ ID No 20099; 654bp; English.
XX
The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
```



```
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1: SEQ ID 8284; 71pp + CD-NOM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 622 BP; 149 A; 126 C; 159 G; 183 T; 5 other:

alignment_scores:
    Quality: 75.50      Length: 73
    Ratio: 1.678        Gaps: 5
    Percent Similarity: 61.644    Percent Identity: 30.137

alignment_block:
US-09-471-276-831 x AAC04209/rev ..

Align seg 1/1 to reverse of: AAC04209 from: 1 to: 622

54 GYASPSerArgLeuProAlaValGlnGluTrpGly...AlaGlnLupr 69
   ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 GCACATACCTACGCTGCTGAATCAACAGTTTGAAGCGTCGACGACGA 273
   69 ovaHlsLeuASerProAlaIlelySHsGlnPheLeuLeuThGlyA 86
   ||||| |||||:|||||:|||||:|||||:|||||:|||||:
272 GCATCACTTA...AGTCCAGAGCTCA.....G 247
   86 sPthrcInclYArgTyFArgCyArGSerClYLeuSerThrclyTrp*** 102
   ||| :|||||:|||||:|||||:|||||:|||||:|||||:
246 ACCAGCGCTGGAAACAGTGAAGCCTCGCTCTACCTCGTCGTCGCTC 197
   103 GlnLeuSerLyLeuLeuGluLeuThrclyProLySValLeu...AlaCy 118
   :|||: |||||:|||||:|||||:|||||:|||||:|||||:
196 CGCTCTC.....CTGACATGCGCTCAGCTGATTACAAAGCTG 162
   118 sSerLeuAlaLeuASpGly 124
   |||||:|||||:|||||:|||||:|||||:|||||:
161 CGCAGCTGACCGCTGATGCG 143

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ90654

seq_documentation_block:
ID AAZ90654 standard: DNA; 735 BP.
XX
AC AAZ90654;
XX
DT 13-JUN-2000 (first entry)
XX
DE Nostoc NspI restriction endonuclease (NspIR) encoding DNA.
XX
KW NspI restriction endonuclease; NspIR; Nostoc; NspI methylase; NspIm; ds.
XX
OS Nostoc sp.
XX
PM US6027929-A.
XX
PD 22-FEB-2000.
```

```
XX
PF 18-AUG-1998; 98US-0135782.
XX
PR 18-AUG-1998; 98US-0135782.
XX
PA (NEME ) NEW ENGLAND BIOLABS INC.
XX
PI Xu S;
XX
DR WPI: 2000-316238/27.
DR P-PSDB: AAY57323.
XX
PT Isolated DNA encoding the NspI endonuclease and methylase, useful for
PT creating recombinant DNA molecules -
XX
PS Claim 1: Fig 3; 14pp; English.
XX
CC The invention relates to a NspI restriction endonuclease (NspIR),
CC obtained from Nostoc sp. The NspIR is recombinantly obtained using a
CC NspI methylase (Nspim) premodified E. coli K strain RRI (gammaDE3).
CC The endonuclease and methylase are useful for creating recombinant DNA
CC molecules. The present sequence represents a DNA encoding a Nostoc NspIR.
XX
SQ Sequence 735 BP; 258 A; 118 C; 169 G; 190 T; 0 other:

alignment_scores:
    Quality: 75.50      Length: 56
    Ratio: 2.097        Gaps: 4
    Percent Similarity: 64.286    Percent Identity: 39.286

alignment_block:
US-09-471-276-831 x AAZ90654 ..

Align seg 1/1 to: AAZ90654 from: 1 to: 735

55 ASPSerArgLeuProAlaValGlnGluTrpGlyAlaGlnLuproValH 71
   ||| |||||:|||||:|||||:|||||:|||||:|||||:
394 GATCATAGATTTCCG...ATGAAAGATGGGAGCTAGAGAGCTCCACA 440
   71 sLeuASer.....ProAlaIlelySHsGlnPhe...LeuLeuT 84
   ||||| |||||:|||||:|||||:|||||:|||||:
441 CPTAACATCTATGATGAATGAATCAAGCAAGTTTCAATTGTTGA 490
   84 hrcGlySPthrcInclYArgTyFArg.....CysArg 94
   |||||:|||||:|||||:|||||:|||||:|||||:
491 AGAAGATATACCTTGCCATGACATCTTTAAATCAAGAACCTGTAG 540
   95 SerGlyLeuSerThrcly 100
   :|||: |||||:|||||:|||||:|||||:|||||:
541 CGATGTATCAAAACTGCT 558

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH85010

seq_documentation_block:
ID AAX85010 standard: DNA; 1863 BP.
XX
AC AAX85010;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human secreted protein gene NO. 78.
XX
DE Human: secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
```

```

XX PN MO9924836-A1.
XX PD
XX PD 20-MAY-1999.
XX PF
XX PF 04-NOV-1998; 98WO-US23435.
XX PR 17-NOV-1997; 97US-0066100.
XX PR 07-NOV-1997; 97US-0064900.
XX PR 07-NOV-1997; 97US-0064908.
XX PR 07-NOV-1997; 97US-0064911.
XX PR 07-NOV-1997; 97US-0064912.
XX PR 07-NOV-1997; 97US-0064983.
XX PR 07-NOV-1997; 97US-0064984.
XX PR 07-NOV-1997; 97US-0064985.
XX PR 07-NOV-1997; 97US-0064987.
XX PR 07-NOV-1997; 97US-0064988.
XX PR 17-NOV-1997; 97US-0066090.
XX PR 17-NOV-1997; 97US-0066094.
XX PR 17-NOV-1997; 97US-0066095.
XX PR 17-NOV-1997; 97US-0066089.
XX PA
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Carter KC, Edner R, Endress GA, Feng P, Janat F;
XX PI Kiyaw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
XX PI Ruben SM, Shi Y, Soppet DR, Wei Y;
XX PI
XX PI WPI: 1999-337740/28.
XX DR P-PSDB: AAY27644, AAY27787, AAY27788, AAY27789, AAY27790, AAY27791,
XX DR AAY27792, AAY27793, AAY27794.
XX PT
XX PT New human secreted proteins and coding sequences useful for treating
XX PT disorders of the immune system and hyperproliferative disorders
XX PS
XX PS Claim 1; Page 322-323; 507pp; English.
XX CC This sequence represents a nucleic acid molecule which encodes a
XX CC secreted human protein. The gene number is given in the descriptor line.
XX CC The gene can be used to generate fusion proteins by linking to the gene
XX CC to a human immunoglobulin Fc portion (e.g. AAX84924) for increasing the
XX CC stability of the fused protein as compared to the human protein only.
XX CC The invention relates to 125 novel genes and their fragments (nucleic
XX CC acid sequences: AAX84933-X85057; amino acid sequences AAY27567-Y27933)
XX CC which are useful for preventing, treating or ameliorating medical
XX CC conditions e.g. by protein or gene therapy. Also, pathological
XX CC conditions can be diagnosed by determining the amount of the new
XX CC polypeptides in a sample or by determining the presence of mutations in
XX CC the new polynucleotides. Specific uses are described for each of the 125
XX CC polynucleotides, based on which tissues they are most highly expressed in
XX CC (see AAX84933 for described uses).
XX SQ
XX SQ Sequence 1863 BP; 457 A; 462 C; 488 G; 450 T; 6 other;

alignment_scores:
    Quality: 75.50      Length: 90
    Ratio: 1.641      Gaps: 5
    Percent Similarity: 51.111      Percent Identity: 27.778

alignment_block:
    US-09-471-276-831 x AAX85010  ..

Align seg 1/1 to: AAX85010 from: 1 to: 1863

4 LeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyProValThrG1 20
      ::::::::::::::::::::
433 ATGGCTGTGAATTTCTGTGGTGGGATGACCTGATCCCGACGGAGAA 482
20 uuaiaaiaailephetryclunhrgin***Seruu.....Trpaaiaclus 35
      :::::      :::::      :::::      :::::      :::::
483 AGCGTCGTCCTGGACACACCTCTCTGCTGACGAGCGGGGCTGGAC 532

```

```

35 eGlnHis**LeuysThrIleu..... 42
    ||| .....|||
533 CAACAGATCTGCCTACTCATTATGTGGTGGAGACATCACTGCT 582
43 .....gYgIn.....cYAsPaLaasPValPr 50
    |||:::
583 ATGAAGAGATCTCGAGGCAAGCTCGTTATTCGGCTCGACACGAGCTGCC 632
50 oGlyProProGlyAspSerArgLeuProAlaValGln.GluTrpGly... 65
   :::: ||||||::::: |||::: |||
633 CTCCTACACGACGAGAAGTCCGCGCTTCCTCGTAATAATGAATGAGCGCTTAA 682
66 AlaGlnGluProValHis 71
   |||:::
683 CTCACGCAATGAATTCAT 700

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT:AAx88978

seq_documentation_block:
ID      AAx88978 standard; DNA; 1906 BP.
XX
XX      AAx88978;
XX
XX      16-SEP-1999 (first entry)
XX
DE      Mouse dephosphorylase inhibiting p91-like protein #4 encoding DNA.
XX
XX      Dephosphorylase inhibiting protein; p91; tyrosine phosphatase SHP-1;
KW      SHP-2; Inositol-5-phosphate SHIP; phosphoryating tyrosine;
KW      immunoreceptor; immunomodulatory agent; ss.
XX
OS      Mus sp.
XX
XX      JP11169184-A.
XX
XX      29-JUN-1999.
XX
XX      12-DEC-1997; 97JP-0362285.
XX
XX      12-DEC-1997; 97JP-0362285.
XX
PA      (UYOK-) UNIV OKAYAMA.
XX
DR      WPI; 1999-422622/36.
XX      P-PDSB; AAY24321.
XX
PT      New peptide - useful for inhibiting dephosphorylase
PS
PS      Claim 7; Page 21-22; 30pp; Japanese.
XX
XX      The present invention describes new proteins for inhibiting
XX      dephosphorylase. The proteins can be combined with tyrosine phosphatase
XX      SHP-1, SHP-2 or inositol-5-phosphatase SHIP by phosphoryating tyrosine.
XX      The present invention also describes an immunoreceptor comprising one
XX      Cn above proteins, and DNA coding the above proteins. The new proteins
XX      can be used as an immunomodulatory agent. The present sequence encodes
XX      a protein from the present invention.
SO      Sequence 1906 BP; 518 A; 546 C; 419 G; 423 T; 0 other;

alignment_scores:
Quality: 75.50 Length: 114
Ratio: 1.302 Gaps: 5
Percent Similarity: 50.877 Percent Identity: 28.947

alignment_block:
US-09-471-276-831 x AAx88978 ..

Align seg 1/1 to: AAx88978 from: 1 to: 1906
||| |||||||::: |||
6 ValPheLeuLeuLeuTrpGlyValThrTTPGlyProValThrluaA.. 21
||| |||||||::: |||
```

```
643 GTCAGCTCTGCTGCTAGCTAATCTCCAAAACCAACCATCAAGCTGA 692
22 .....AlaIlePheTyrGluThrGln**SerLeuTPraIaGluS 35
693 ACCAGGATGTGTGATGCGCTCCAAAGACCAATGACCATCTGTGTGACG 741
35 ergIuhIs**LeuIysThrLeuGluGlnCysAspAlaAspVal..... 49
742 .....GGCAACCTGTGATCAGACAGATATTTT 768
50 ProGlyProProGlyAspSerArgLeuProAlaValaGluThrGlyAl 66
769 CTCGCAATATGACGGAAGCCAAAACACACAGACACAGACC..... 810
66 agInGluProValHisLeuAspSerProAlaIleLysHisGlnPheLeu 83
811 .....CTACAGCAGCGCTGGGACAAAGCGCATGCTTTTCA 844
83 euThrGlyAspThrGln.....GlyArgTyrArgCysArgSerGly 96
845 TTCCTTCTGTGACACAAACAAATGACGGCAATATGCTGTATTGTTAC 894
97 LeuSerThrGlyTyr**GlnLeuSerLysLeuLeuGluLeu 110
895 ACCTCAGCTGCTGTGTCACAGCCCGCTGAGCTG 936

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAF18310

seq_documentation_block:
ID  AAF18310 standard: DNA; 2373 BP.
XX
XX  AAF18310;
XX
XX  14-MAR-2001 (first entry)
XX
DE  Lung cancer associated polynucleotide sequence SEQ ID 329.
XX
KW  Human; lung cancer associated protein; neuroprotective; cytostatic;
KW  cardioclitive; immunomodulatory; muscular active; vulnerary;
KW  gastrointestinal; nephrotropic; antinfective; gynecological;
KW  antibacteriati; diagnosis; neural disorder; immune disorder; reproductive;
KW  proliferative disorder; wound healing; infectious disease; ds.
XX
OS  Homo sapiens.
XX
XX  WO20005180-A2.
XX
XX  21-SEP-2000.
XX
XX  08-MAR-2000; 2000WO-US05918.
XX
XX  12-MAR-1999; 99US-0124270.
XX
XX  (HUMA-) HUMAN GENOME SCT INC.
XX  (ROSE/) ROSEN C A.
XX
XX  Ruben SM;
XX
XX  WPI: 2000-587514/55.
XX  P-PSDB: AAB58434.
XX
XX  Lung cancer associated gene sequences, referred to as lung cancer
XX  PT  antigens, useful for treatment, prevention, and diagnosis of disorders
XX  PT  such as lung cancer -
XX
XX  Claim 1: Page 786-787; 1425pp: English.
XX
XX  Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
XX  CC  associated proteins represented in AAB58106 - AAB58548. Lung cancer
XX  CC  associated proteins and polynucleotide sequences, their agonists, and
XX  CC  antagonists may have neuroprotective; cytostatic; cardioclitive;
XX  CC  immunomodulatory; muscular active general; vulnerary; gastrointestinal
XX  CC  general; nephrotropic; antinfective; gynecological; or antibacterial
```

```
CC  activity. The invention also includes antibodies specific for the
CC  protein or polynucleotide sequences. The lung cancer associated
CC  polynucleotide sequences may be used for detection of lung cancer,
CC  CC  chromosome identification, as chromosome markers, and for numerous other
CC  CC  diagnostic or research purposes. The proteins may be used to treat
CC  CC  disorders such as neural, immune, muscular, reproductive,
CC  CC  gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC  CC  disorders. The proteins may also be used in the treatment of wounds and
CC  CC  infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC  CC  peptide AAB58549 are used in the course of the invention for the
CC  CC  identification and characterisation of the polynucleotide and protein
CC  CC  sequences.
XX
XX  Sequence 2373 BP; 456 A; 730 C; 669 G; 514 T; 4 other:
XX
XX  alignment_scores:
XX      Quality: 75.50      Length: 90
XX      Ratio: 1.641      Caps: 5
XX      Percent Similarity: 51.111      Percent Identity: 27.778
XX
XX  alignment_block:
XX  US-09-471-276-831 x AAF18310 ..
XX
XX  Align seg 1/1 to: AAF18310 from: 1 to: 2373
XX
XX  4 LeuValValPheLeuLeuLeuTyrGlyValThrTrpGlyProValThrGlu 20
XX  ::::||||| :||| |||||:||||| |||
XX  733 ATGCGTGTAAATTTCTGTGCGGATGACCTGGATCCCGACGGGAA 782
XX  20 uAlaAlaIlePheTyrGluThrGln**SerLeu.....TrpAlaGluS 35
XX  :||||: :|: :|: ||| :|||:|:
XX  783 AGCTGCTCCTGTGACACAACTCTGTGACAGCGGGGCTGAGCTGAC 832
XX  35 ergIuhIs**LeuIysThrLeu..... 42
XX  ||| :|||
XX  833 CAACACATCTGCTTACCTACTATGCTGCGGTGAGACATCACTGTGT 882
XX  43 .....GlyGln.....CysAspAlaAspValPr 50
XX  883 ATGAAGATGCTGAGGCGAGCTGTTATCGCTGCGACAGGAGCTGCG 932
XX  50 oGlyProProGlyAspSerArgLeuProAlaValaGln.GluTyrGly... 65
XX  :||: |||||:|: :|||:|: :|||:|:
XX  933 CTCGTACACCGAAGTTCGGCTTTCCTGCTCGGAATGATATGCGCGTTA 982
XX  66 AlaGlnGluProValHis 71
XX  |||: :|||
XX  983 CTCACGCAATGAATTCAT 1000
```

OM of: US-09-471-276-831 to: GenEmbl:* out_format: pfs
Date: Jan 7, 2002 5:43 PM

About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framed_p2n_model -DEV=x1h
-O=/cgn2_1/USPTO_pool/US09471276/runat_07012002_142445_19882/app_query.fasta.1.274
-DB=GenEmbl -GPM=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELLOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blomsme2 -TRANS=humana0.cdi
-LIST=100 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09471276@CGN1_1.6345
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-471-276-831
Query length: 126
Database: GenEmbl:*
Database sequences: 1472140
Search time (sec): 1351.940000

score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_hhg:AC012313	-	457.00	1.1e-36	186451	AC012313 Homo sapiens chr10
gb_pr:AC027222	-	222.50	4.9e-14	2175	AK027222 Homo sapiens cDNA: FL
gb_pat:AX045627	+	139.50	243.96	1908	AX045627 Sequence 7 from Paten
gb_ro:RNO302031	+	132.50	235.52	1741	AJ302031 Rattus norvegicus mRN
gb_pat:AX046966	+	108.00	192.01	956	AX046966 Sequence 33 from Paten
gb_pr:AF324830	+	108.00	191.56	1016	AX046966 Sequence 29 from Paten
gb_pr:HS091925	+	103.50	179.76	1603	AF324830 Homo sapiens Immunogl
gb_pr:AF025532	+	103.50	179.66	1625	U91925 Human clone HM18 monoc
gb_pr:AF025532	+	103.50	179.66	1625	AF025532 Homo sapiens monocyt
gb_pr:AF009007	+	103.00	177.13	1593	AF009007 Homo sapiens leucocyt
gb_pr:AF009005	+	103.00	176.85	2094	AF009005 Homo sapiens immunogl
gb_pat:AX046940	+	103.00	174.46	2777	AX046940 Sequence 3 from Paten
gb_pr:AF004230	+	103.00	169.93	2897	AF004230 Homo sapiens monocyt
gb_pr:AF072099	+	103.00	166.21	8872	AF072099 Homo sapiens leucocyt
gb_pr:AF089768	+	103.00	166.21	47084	AF089768 Homo sapiens leucocyt
gb_pr:AC011515	+	103.00	163.90	148497	AC011515 Homo sapiens chr10
gb_hhg:AL139117	+	102.50	178.91	1386	AL139117 Homo sapiens chr10
gb_pr:HS082979	+	102.50	153.26	39829	U82979 Human Immunoglobulin-1
gb_pr:AC008884	+	102.00	146.69	97137	AC008884 Homo sapiens chr10
gb_hhg:AL355172	+	98.50	168.93	1953	AL355172 Homo sapiens chr10
gb_pr:AF009220	+	98.50	168.59	2046	AF009220 Homo sapiens leucocyt
gb_pr:AF009006	+	98.50	168.42	2094	AF009006 Homo sapiens immunogl
gb_pr:HS082279	+	98.50	166.30	2790	HS082279 Human Immunoglobulin-1
gb_pat:AX046938	+	96.00	165.96	2922	AX046938 Sequence 21 from Paten
gb_pr:AC010492	+	96.00	163.91	84087	AC010492 Homo sapiens chr10
gb_pr:AF041262	+	94.50	163.91	1395	AF041262 Homo sapiens chr10
gb_ro:AF322217	+	93.50	159.14	1745	AF322217 Rattus norvegicus inh
gb_pr:AF031554	+	93.50	157.78	2066	AF031554 Homo sapiens chr10
gb_ro:AF322216	+	92.50	157.78	1407	AF322216 Rattus norvegicus inh
gb_pr:AF009633	+	92.50	157.28	2063	AF009633 Homo sapiens chr10
gb_pr:AF009633	+	92.50	157.28	2063	AF009633 Homo sapiens chr10
gb_pr:AF031555	+	92.50	157.28	2063	AF031555 Homo sapiens chr10
gb_pr:AF009635	+	92.50	157.27	2066	AF009635 Homo sapiens chr10
gb_pr:AF009641	+	92.50	157.27	2066	AF009641 Homo sapiens chr10
gb_pr:AF009642	+	92.50	157.27	2066	AF009642 Homo sapiens chr10
gb_pr:AF009636	+	92.50	157.18	2091	AF009636 Homo sapiens chr10

gb_pr:AF009640	+	92.50	157.18	1.77	2091	AF009640 Homo sapiens chr10
gb_pr:AC005590	-	92.50	151.12	3.85	4756	AC005590 Homo sapiens chr10
gb_hhg:AC041031	-	91.00	128.67	68.50	68207	AC041031 Homo sapiens chr10
gb_pat:AX046972	+	90.50	158.89	1.42	997	AX046972 Sequence 35 from Pat
gb_pr:AF009644	+	90.50	158.82	1.43	1007	AX046968 Sequence 31 from P
gb_pr:AF0030575	+	90.50	153.54	2.82	1983	AF009644 Homo sapiens chr10
gb_pr:AF031556	+	90.50	153.53	2.83	2061	AF0030575 Homo sapiens chr10
gb_pr:AF009637	+	90.50	153.35	2.89	2063	AF031556 Homo sapiens chr10
gb_pat:AX046956	+	90.50	153.08	3.00	2114	AF009637 Homo sapiens chr10
gb_pr:AF025533	+	90.50	153.08	3.00	2194	AX046956 Sequence 19 from P
gb_pr:AF025533	+	90.50	153.08	3.00	2194	AF025533 Homo sapiens chr10
gb_hhg:AC011124	+	90.50	129.07	65.08	56896	AC011124 Homo sapiens chr10
gb_hhg:AC0118933	+	90.50	120.40	197.94	144489	AC018933 Homo sapiens chr10
gb_pr:AC009638	-	90.50	119.61	219.00	205307	AC009638 Homo sapiens chr10
gb_pr:AF009638	+	89.50	151.66	3.59	2063	AF009638 Homo sapiens chr10
gb_pr:AF025529	+	89.50	150.40	4.22	2445	AF025529 Homo sapiens chr10
gb_pat:AX046948	+	89.50	150.40	4.22	2446	AX046948 Sequence 11 from P
gb_hhg:AC0118933	+	88.50	117.92	272.06	155313	AC006453 Homo sapiens chr10
gb_pr:AC027612	-	88.50	117.79	276.60	158052	AC027612 Homo sapiens chr10
gb_hhg:AC012314	-	88.00	122.16	157.96	159503	AC012314 Homo sapiens chr10
gb_hhg:AC032037	-	87.50	152.71	3.14	75966	AC032037 Homo sapiens chr10
gb_pr:AF012842	+	87.50	147.91	5.81	1075	AF012842 Homo sapiens chr10
gb_pr:AF009633	+	87.50	147.91	5.81	2063	AF009633 Homo sapiens chr10
gb_pr:AF009634	+	87.50	147.81	5.89	2091	AF009634 Homo sapiens chr10
gb_pr:HS091928	+	87.50	147.53	6.10	2170	U91928 Human clone H19 mono
gb_hhg:AC055877	+	87.50	114.86	402.84	182414	AC055877 Homo sapiens chr10
gb_hhg:AC016888	+	87.00	113.85	458.70	184284	AC016888 Homo sapiens chr10
gb_pr:AF189277	+	86.50	136.45	25.27	7568	AF189277 Homo sapiens chr10
gb_pr:AC009892	+	86.50	113.26	382.74	134008	AC009892 Homo sapiens chr10
gb_ro:AF289664	-	86.50	113.87	457.54	161852	AF289664 Mus musculus chr10
gb_ro:AF289667	-	86.50	112.25	563.16	201605	AF289667 Mus musculus chr10
gb_hhg:AC009456	-	86.00	115.99	348.49	106873	AC009456 Homo sapiens chr10
gb_hhg:AL355987	-	86.00	112.06	377.07	182179	AL355987 Homo sapiens chr10
gb_hhg:AL449425	-	86.00	111.56	614.92	194840	AL449425 Homo sapiens chr10
gb_ba:SM591789	+	86.00	108.51	909.69	294800	AL591789 Simorhizobium me
gb_pr:AC009080	+	85.50	110.44	710.46	199891	AL079903 Human chromosome
gb_ba:MSB937CS	+	85.00	121.57	170.49	38914	L78820 Mycobacterium lepra
gb_pr:AC009123	+	85.00	109.61	799.43	197048	AC009123 Homo sapiens chr10
gb_ba:MLEPRTN2	+	85.00	105.50	1.3e+03	344050	AL583918 Mycobacterium le
gb_pat:AX046950	+	84.50	142.85	11.12	1910	AX046950 Sequence 13 from P
gb_pr:AF025530	+	84.50	142.29	11.95	2061	AF025530 Homo sapiens chr10
gb_pat:AF025534	+	84.50	142.29	11.95	2061	AF025534 Homo sapiens chr10
gb_ro:AF035553	-	84.50	130.49	54.28	10216	AF035553 Mus musculus Na/K
gb_pr:HS24985	-	84.00	110.80	678.22	129968	AL035591 Human DNA sequen
gb_hhg:AC010550	+	84.00	106.82	1.1e+03	229294	AC010550 Homo sapiens chr10
gb_hhg:AC009090	+	84.00	106.40	1.1e+03	226113	AC009090 Homo sapiens chr10
gb_hhg:AC023825	+	83.50	107.90	983.69	169591	AC023825 Homo sapiens chr10
gb_hhg:AC021612	+	83.00	111.62	610.24	90140	AC021612 Homo sapiens chr10
gb_hhg:AC020769	+	83.00	106.71	1.1e+03	175493	AC020769 Homo sapiens chr10
gb_pr:AF034198	+	82.50	132.95	39.59	4401	AF034198 Homo sapiens chr10
gb_pr:AF002362	+	82.50	131.42	48.15	5413	AB002362 Human mRNA for KIA
gb_hhg:AC007082	+	82.50	112.92	516.56	165652	AC007082 Homo sapiens chr10
gb_pr:AC007934	+	82.50	106.56	1.2e+03	165652	AC007934 Homo sapiens chr10
gb_hhg:AF001155	+	82.50	106.29	1.2e+03	163709	AF001155 Homo sapiens chr10
gb_pr:AC074121	+	82.50	105.17	1.2e+03	166379	AC074121 Homo sapiens chr10
gb_hhg:AC025491	+	82.50	105.59	1.3e+03	179336	AC025491 Homo sapiens chr10
em_hhg_hum:AC023091	+	82.50	105.54	1.3e+03	181068	AC023091 Homo sapiens chr10

seq_name: gb_hhg:AC012313

seq_documentation_block:

LOCUS AC012313 186451 bp DNA

DEFINITION Homo sapiens chromosome 19 clone CTD-2619J13, ** SEQUENCING IN

PROGRESS ***, 2 ordered pieces.

ACCESSION AC012313 6 GI:15145671

VERSION AC012313.6

KEYWORDS HTGS:HTGS-PHASE2; HTGS:ACTIVEPTN.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.
 1 (bases 1 to 186451)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 186451)
 2 (bases 1 to 186451)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (23-OCT-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Aug 9, 2001 this sequence version replaced gi:7711546.
 * NOTE: This is a "working draft" sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 181942: contig of 181942 bp in length
 * 181943 182042: gap of unknown length
 * 182043 186451: contig of 4409 bp in length.
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

 Project Information
 Center Project Name: 835906
 Center clone name: CTD-2619J13

 Summary Statistics
 Consensus quality: 184784 bases at least Q40
 Consensus quality: 185394 bases at least Q30
 Consensus quality: 185487 bases at least Q20
 Estimated insert size: 194270; agarose-1p estimation
 Estimated insert size: 185541; sum-of-contigs
 estimation
 Quality coverage: 0.95 in Q20 bases; agarose-1p
 estimation
 Quality coverage: 1 in Q20 bases; sum-of-contigs
 estimation.
 Location/Qualifiers
 1. 186451
 /organism:"Homo sapiens"
 /db_xref:"taxon:9606"
 /chromosome:"19"
 /clone:"CTD-2619J13"
 BASE COUNT 44285 a 49190 c 49575 g 43301 t 100 others
 ORIGIN
 alignment_scores:
 Quality: 457.00 Length: 172
 Ratio: 4.231 Gaps: 5
 Percent Similarity: 62.791 Percent Identity: 62.791
 alignment_block:
 US-09-471-276-831 x AC012313/rev ..
 Align seg 1/1 to reverse of: AC012313 from: 1 to: 186451
 1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGly..... 12
 11643 ATGTCCATGCTGCTGCTTCTTCTTCTTCTGAGGCTCTCGGG 11594
 12 12
 11593 TCTGCGTCTCTGGGCGGAGGCGGGAACAGACCTTACGGGGCT 11544
 13Val..ThrTrpGlyProValThrGluAlaIle..Phe..24
 11544

11543 GTCTCCGACAGTGTACCTGGGCCAGTGACAGAAAGACCATATGTGA 11494
 24 24
 11493 GTCCAGGCGTACGCTGTGGAAATGTGGCCAGCCGGGGAGCCCT 11444
 25TyrGluTrh 27
 11443 GTCTCTGTGACACGTGACCTGATCCAGCCACCCCTTGCACTTATGAGAC 11394
 27 rcin***SerLeuTrpAlaGluSerGluHis***LeuLysThrLeuGly 44
 11393 GCAGCCAGCCCTGTGGCGACAGTCCGAAATCATCTGTGAA..ACCTTGGGCC 11345
 44 IncyAspAlaAspValProGlyProProGlyAspSerArgLeuProAla 60
 11344 AATGTACGCTGACGTGCGCAGGCCCTGTGAAACATCCACACTTCAGCT 11295
 61 ValGlnGlnTrpGly..AlaGlnGlnProValHisLeuAspSerProAla 77
 11294 GTTCAGAAATGGGCGTGGCCAGGAGCCCTGTGCACTTGAATCACTGAC 11245
 77 LeuLysHisGlnPheLeuLeuThrGlyAspThrGlnGlyArgTrpGly 93
 11244 TCAGGACACGATTCCTGCTGACGGGTGACACCCAGGCCCTACCGCTGC 11195
 94 ArgSerGlyLeuSerThrGlyTrp***GlnLeuSerLysLeuGluLe 110
 11194 CGCTCGGCGCTTGTCCACAGATGAGACCCAGCTGAGCAAGCTCTGAGCT 11145
 11144 GACAGGCCCA 11135
 110 utrhrGlyPro 113
 seq_name: gb.pr:AK027222
 seq_documentation_block:
 LOCUS AK027222 2175 bp mRNA PRI 29-SEP-2000
 DEFINITION Homo sapiens cDNA: FLJ23569 fls, clone LNG11861.
 ACCESSION AK027222
 VERSION AK027222.1 GI:10440301
 KEYWORDS oligo capping; fls (full insert sequence).
 SOURCE Homo sapiens human lung cDNA to mRNA, clone_11b: LNG clone: LNG11861.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.
 REFERENCE 1 (sites)
 Kawakami,T., Noguchi,S., Itoh,T., Shigetani,K., Senba,T.,
 Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A.,
 Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hirao,M.,
 Ohmori,Y., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T.,
 Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
 NEDO human cDNA sequencing project
 TITLE Unpublished (2000)
 JOURNAL 2 (bases 1 to 2175)
 REFERENCE Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
 Shibahara,T., Tanaka,T. and Nakamura,Y.
 JOURNAL Direct Submission
 Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
 Sugano, Institute of Medical Science, University of Tokyo,
 Laboratory of Genome Structure Analysis, Human Genome Center,
 Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan
 (E-mail:cdna@elms.u-tokyo.ac.jp, Tel:81-3-5449-5286,
 Fax:81-3-5449-5416)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan: cDNA full insert
 sequencing: Research Association for Biotechnology: cDNA library
 construction, 5'- & 3'-end one pass sequencing: Department of
 Virology and Human Genome Center, Institute of Medical Science,
 University of Tokyo (partly supported by Science and Technology
 Agency).
 FEATURES
 source 1..2175
 Location/Qualifiers

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="LNC11861"
/clone_lib="LNC"
/tissue_type="human lung"
/note="cloning vector pME18SFL3"

BASE COUNT      520 a      598 c      626 g      431 t
ORIGIN

alignment_scores:
    Quality: 222.50      Length: 115
    Ratio: 3.973      Gaps: 3
    Percent Similarity: 48.696      Percent Identity: 46.957

alignment_block:
US-09-471-276-831 x AK027222/rev ..
Align seg 1/1 to reverse of: AK027222 from: 1 to: 2175

1 MetSerMetLeuValValPheLeuLeuLeuTrpGly..ValThrTrpGly 16
|||||
1134 ATGTCCATGCTGCTGGTCTTCTCTGCTGCTGAGGCTCTGCTGGG 1085
16 YPro.ValThrGluAlaAlaIlePheTyrgluThrGln**SerLeuTrp 32
|||
1084 TCTGGGTC..... 1077
33 AlaGluSerGluHis**LeuLysThrLeuGlyGlnCysAspAlaAspVal 49
1077 ..... 1077
49 lProGlyProProGlyAspSerArgLeuProAlaValGlnGluTrpGlyVal 66
1077 ..... 1077
66 lAcGlnGluProValHisLeuAspSerProAlaIleLysHisGlnPheLeu 82
|||||
1076 .....CCTGCCATTCACAGCACCACTTCTCTG 1053
83 LeuThrGlyAspThrGlnGlyArgTyArgCysArgSerGlyLeuSerThr 99
|||||
1052 CTGACGGGTGACACCCAGGCGCCCTACCGCTGCGCGCTGCGGCTGTCCAC 1003
99 rGlyTrp**GlnLeuSerLysLeuLeuGlnLeuThrGlyPro 113
|||||
1002 AGCATGGAGCCAGCTGAGCAGCTCTGAGACTGACAGGCGCA 960

seq_name: gb_pat:AX045627

seq_documentation_block:
LOCUS AX045627 1908 bp DNA PAT 24-NOV-2000
DEFINITION Sequence 7 from Patent WO0066787.
ACCESSION AX045627
VERSION AX045627.1 GI:11344029
KEYWORDS
SOURCE
ORGANISM house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 1908)
Kopchick,J.J. and Tjong,J.
Growth hormone-regulatable liver genes and proteins, and uses
thereof
Patent: WO 0066787-A 7 09-NOV-2000;
Ohio University (US)
LOCATION/Qualifiers
1..1908
/organism="Mus musculus"
/db_xref="taxon:10090"

BASE COUNT      525 a      475 c      477 g      431 t
ORIGIN
```

```
alignment_scores:
    Quality: 139.50      Length: 138
    Ratio: 1.744      Gaps: 4
    Percent Similarity: 57.971      Percent Identity: 29.710

alignment_block:
US-09-471-276-831 x AX045627 ..
Align seg 1/1 to: AX045627 from: 1 to: 1908

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyTrp 17
|||||
11 ATGTCTCTGCTGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
17 oValThrGluAlaAlaIlePheTyrgluThrGln**SerLeuTrpAlaG 34
|||
61 AGGAATACCTCTTAATGCTCGATCTGCGCAGTGAACCTTAATGCTGCGCAG 110
34 lUserGluHis**LeuLysThrLeuGlyGln.....CysAsp 46
|||
111 AGCCTGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 160
47 AlaAspValProGlyProProGlyAspSerArgLeuProAlaValGlnGly 63
|||
161 GTTCATTTGCGC.....ACTAAGCTCTTCGACGTCGATCCAGAA 198
63 uTrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleLysHisG 80
|||
199 CGGGTGCTTCCTGAGTCAGTCCGACCTTGACACAGCTGCTGCTCATPACC 248
80 lnPheLeuLeu.....ThrGlyAspThrGlnGlyArgTyArgCys 93
|||
249 GCTTTTCCCTGGGGCCATTACAAAGTACAAACAGTGCATCTACCGCTGC 298
94 ArgSerGlyLeuSer.....ThrGly 100
|||
299 AGATGTGGCGCTGGAACCCCTGTTGACATTCACCTGCCAGCAGTGAACAA 348
100 YTrp**GlnLeuSerLysLeuLeuGlnLeuThrGlyProLysValLeu 117
|||||
349 GTGAGACCATGCTTAAGCATTCCTGTGAGAGTGAACAGGAAGACGCCCTTGC 398
117 lAcYSerLeuAla 121
|||
399 CTGCGCCCTTGGCT 412

seq_name: gb_ro:RNO302031

seq_documentation_block:
LOCUS RNO302031 1741 bp mRNA ROD 03-JUL-2001
DEFINITION Rattus norvegicus mRNA for putative alpha 1B-glycoprotein (ORF1).
ACCESSION AJ302031
VERSION AJ302031.1 GI:11877347
KEYWORDS alpha 1B-glycoprotein.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1741)
Gardmo,C., Persson,B. and Mode,A.
Cloning of a novel growth hormone-regulated rat complementary
deoxyribonucleic acid with homology to the human
alpha1B-glycoprotein, characterizing a new protein family
Endocrinology 142 (6), 2695-2701 (2001)
21255654
JOURNAL
MEDLINE
2 (bases 1 to 1741)
Gardmo,C.
REFERENCE
Direct Submission
AUTHORS
Submitted (12-DEC-2000) Gardmo C., Department of Medical Nutrition,
Karolinska Institutet, Novum F60, Huddinge, S-14186, SWEDEN
JOURNAL
LOCATION/Qualifiers
```

```

source
1.1741
/organism:"Rattus norvegicus"
/strain:"Sprague Dawley"
/db_xref:"taxon:10116"
/sex:"female"
/tissue_type:"liver"
/dev_stage:"adult"
30.1571
/note:"ORF1"
/codon_start:1
/product:"putative alpha 1B-glycoprotein"
/protein_id:"CAC19029.1"
/db_xref:"GI:11877348"
/db_xref:"SPTREMBL:Q9EPH1"
/translation:"MSLLTIVLLGFLTGPGNALNLNSGSEPELRABEPOSULEPMAN
LTVCAVDLPKRVFELINMGWFLISOVLRYPSLIGATTSNSGVRCGVEP
PYDIOPLALSKRTMLSNLEVTGKPEPLPPSAHADPVMHTTGGILPYIMCVARGV
TYLLRKEGVDTQKPDVQHKGTAGFLIKPCNVSCSYLTHAGGEPSESAIVT1KMSA
TOLPSPCLIMCSYLTIPQKTHETLACKAPRAAEFOLROGRLVNIQGFPTTRDATI
YVNLKELDNOSPTFCRYRMKHYMHVMSDESPVLEMSDELPAPVLTAEPSHNL
PGSTVOLRCTAHKAGLRGLOROCKPDLVYVOMLNSGTEAVFELINISTDSGNSC
LYMEQAPPPSGSASSELELRINGPAPRLEALMKGVPLDHEAIFQCHGVPRVSM
ELVRGEFTPFMASTSTSAFLKLFVGPQHTGNVSCRVTALSPFTESGISDVEV
VEGS"
99
/note:"translation change from Asn to Asp"
/replace:"g"
1692
BASE COUNT      450 a      458 c      430 g      403 t
ORIGIN

alignment_scores:
Quality: 132.50      Length: 166
Ratio: 1.523      Gaps: 6
Percent Similarity: 52.410      Percent Identity: 28.313

alignment_block:
us-09-471-276-831 x RNO302031 ..
Align seq 1/1 to: RNO302031 from: 1 to: 1741

1 MetSerMetLeuValValPheLeuLeuTrpGlyValThrTrpGlyPr 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30 ATGCTCTGCTGTTGACTGACTGACTCTCTCGGGCTTCACTCGGGCCC 79
17 oValThGluAlaAlaIlePheTyrGluThrGln**SerLeuTrpAlaG 34
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
80 AGCAAAATGCTCTTGGCTTAATTCGGCACTGAACCTGAACCTACGGCAG 129
34 luserGluHis**LeuLysThrLeuGlyGln.....CysAsp 46
||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
130 AGCCTCAATGCTTGGCTGGAACCTGGCAAACTCAGACCTGCTGCGCA 179
47 AlaAspValProGlyProProGlyAspSerArgLeuProAlaValAlaGlnG 63
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
180 GTGATGTTGGCTAC.....CAAGCTTTCGAGCTGATCATGA 216
63 uTrpGlyAlaGlnGlnProValHisLeuAspSerProAlaIleLysHis 79
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
217 ATGGCTGTTGCTTCAGTCAGCTCGAGTGAACACCGGCTCTTCATAT 266
80 GluPheLeuLeu.....ThrGlyAspThrGlnGlyArgTyrArgGly 93
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
267 CGCTTTTCCTGGGGCCATTAACAAGTACAGACGCTTACCGCTG 316
93 sArgSerGlyLeuSer.....:::
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
317 CCGATGTGGGGGTGGAAACCCCTGTTGACATTCAACCTCCAGCGCTGACGA 366
100 lYTrp**GlnLeuSerLysLeuLeuGlnLeuThrGly.....112
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
367 AGTGCACAATGCTTAACCAATGCTTTGAGGTGACAGGGAAGACCCCTTG 416

```

```

112 .....112
417 CCTCCACCCCTCAGCTCAGCTGATCCAGTCTCCTGGATCAGACTGGTG 466
113 ....ProLysValLeuAlaCysSerLeuAlaLeuAspGlyAlaSer 126
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
467 CCTGCTGCTCATGATGTGCGGGCTTGCCAGTGCAGGCTGAAC 512
seq_name: gb_pat:AX046970
seq_documentation_block:
LOCUS      AX046970      956 bp      DNA
DEFINITION Sequence 33 from Patent WO068383.
ACCESSION      AX046970
VERSION      AX046970.1 GI:11876399
KEYWORDS
SOURCE
ORGANISM      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 956)
AUTHORS      Cosman,D.D., Anderson,D.M. and Borges,L.
TITLE      Family of immunoregulators designated leukocyte immunoglobulin-1like
receptors (Lir)
JOURNAL      Patent: WO 0068383-A 33 16-NOV-2000;
IMMUNEX CORPORATION (US)
FEATURES
source      Location/Qualifiers
CDS
1..956
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
115..912
/note:"unnamed protein product"
/codon_start:1
/protein_id:"CAC18938.1"
/db_xref:"GI:11876400"
/translation:"MAPSHPSAQLOPVGDAVSPALNVLCLGSLGPRTHVQAGNL
SKATLMAEPGSYISRNSVTI RCQGTLEAOEYLKRGSPRPMDTONLEPKNKAIFS
IPSMETHAGRCYRYVSPAGMSPEPDLLELVYGFYKPTLSALSPVYVSGENVTL
OCGRRLRDRLITREGBDKLSWTLDSOLTFRSGQOALFPVGPATPSRMALRCTGSR
RHLDVMSPEPDLLEIPVSGEATVVFSTIGOSYTGCELYRGSPC"
BASE COUNT      220 a      291 c      269 g      176 t
ORIGIN

alignment_scores:
Quality: 108.00      Length: 120
Ratio: 1.612      Gaps: 6
Percent Similarity: 55.833      Percent Identity: 34.167

alignment_block:
us-09-471-276-831 x AX046970 ..
Align seq 1/1 to: AX046970 from: 1 to: 956

1 MetSerMetLeuValValPheLeuLeuTrpGlyValThrTrpGlyPr 17
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
169 GTGAGCCCTGCTCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 218
17 oValThGluAlaAlaIlePheTyrGluThrGln**SerLeuTrpAlaG 34
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
219 CAGGACCCAGCTGACGAGCAACCTTCACAAAGCCACCTCTGGGCTG 268
34 luserGluHis**LeuLysThrLeuGlyGln..CysAspAlaAspValP 50
||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
269 AGCCAGGCTCTGTCATC...AGCCGCGGCAACTCTGACCATCCGCTG 315
50 roGlyProProGlyAspSerArgLeuProAlaValGlnGln.....63
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
316 CAGGAGACCCCTGGAGGCCAGCAATACCTGCTGTTAAAGAGGAGACCC 365
64 .....TrpGlyAlaGlnGlnProValHisLeuAspSerProAlaIleL 78
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```



```

366 AGAACCCCTGGACACAGAACCCCACTG.....GAGCCCAAGACAA 406
78 ysh1sglnpheuleuThrglyaspThrgln.....GlyargTyr 91
11 ::::: ::::: ::::: ::::: :::::
407 AGGCAAGATTTCATCCATCCATGACAGACACATGACGAGATAC 456
92 ArgGcArgSerGlyLeuSerThrglyTrrp**GlnLeuSerLysLeuLe 108
457 GCCTGTACTACTACAGCCCTGAGCGTGGTCAGAGCCAGACCCCT 506
108 ugluleu 110
1111111
507 GGAGCTG 513

seq_name: gb_pat:AX046966

seq_documentation_block:
LOCUS AX046966 1016 bp DNA PAT 15-DEC-2000
DEFINITION Sequence 29 from Patent WO0068383.
ACCESSION AX046966
VERSION AX046966.1 GI:11876395
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1016)
Cosman,D.-J., Anderson,D.M. and Borges,L.
Family of immunoregulators designated leukocyte immunoglobulin-like
receptors (LIR)
Patent: WO 0068383-A 29 16-NOV-2000;
IMMUNEX CORPORATION (US)
FEATURES
Source
Location/Qualifiers
1..1016
/organism="Homo sapiens"
/db_xref="taxon:9606"
69..968
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:11876396"
/protein_id="CAC18936.1"
/translation="MAPWSHPSAQLQEVGDPAYSPALMVLICLGLSLGPRTHYQAGNL
SKALMAEPGVSIRGNSVTRCOGTLEADQERYLWKEGSEPMOTQNPLEPKNKARFS
IPSTEHHAAGRCYCYXSPAGMSESPDLLEVFVGNKPTLSALPSPVYTSGENVRL
QCGSRLLRPDRILLFEEDHKLSPTLSQDLPVSGFOALFPGVPYPSHRMLRCYGSRR
RHLYQWSESPDLLETFYSGADNLSSQNKSDSGTASHLDYAVENLRNMGAGLTL
VVLILFLQDMHSRSPQAAAGR."
BASE COUNT 231 a 310 c 282 g 193 t
ORIGIN
alignment_scores:
Quality: 108.00 Length: 120
Ratio: 1.612 Gaps: 6
Percent Similarity: 55.833 Percent Identity: 34.167

alignment_block:
US-09-471-276-831 x AX046966 ..

Align seg 1/1 to: AX046966 from: 1 to: 1016

1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrrpGlyTr 17
::::: ::::: ::::: ::::: :::::
123 GTAGAGCCCTGCCCTCATGCTTCTGCTGCGCTGAGGTGGGCC 172
17 ovalThrgLAlaAlaAlaIlePheTrrGluThrgln***SerLeuTrpAlaG 34
1 ::::: ::::: ::::: ::::: :::::
173 CAGGACCCAGCTGCAGGACGAGAACTCTCAAGACCACTCTGGGCTG 222
34 IuSerGlnHis***LeuLysThrLeuGlyGln..CysAspAlaAspValP 50
11 ::::: ::::: ::::: :::::
223 AGCCAGGCTCTGTGATC...AGCCGGGGGAAACCTGCTGACCATCCGCTG 269

```

```

50 1061ProGlyProGlyAspSerArgLeuProAlaValGlnGlu..... 63
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
270 CAGGGAGACCCCTGGAGCCAGAGAAATACCTCTGCTTAAGAGGAGAGCCC 319
64 .....TPrGlyAlaGlnGluProValHisLeuAspSerProAlaIle 78
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
320 AGAACCCCTGGACACAGACCCACTG.....GAGCCCAAGACAA 360
78 yHisGlnPheLeuLeuThrGlyAspThrGln.....GlyArgTrp 91
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
361 AGCCGAGATTCATCCATCCATGACAGACAGACATGACAGGAGATAC 410
92 ArgGlyAspSerGlyLeuSerThrGlyTrp**GlnLeuSerLysLeuLe 108
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
411 CCTGTTACTACTACAGCCCTGACGCTGTGTACAGCCAGGAGCCCT 460
108 uGluLeu 110
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
461 GGAGCTG 467
seq_name: gb_AF324830

seq_documentation_block:
LOCUS AF324830 1365 bp mRNA PRI 12-MAY-2001
DEFINITION Homo sapiens Immunoglobulin-like transcript 11 protein (ILT11)
ACCESSION AF324830
VERSION AF324830
KEYWORDS AF324830.1 GI:14028659
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1365)
AUTHORS Wende,H., Ziegler,A. and Volz,A.
TITLE Genomic organization of the ILT11 gene, a novel member of the
Leukocyte receptor cluster (LRC)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1365)
AUTHORS Wende,H., Ziegler,A. and Volz,A.
TITLE Direct Submission
JOURNAL Submitted (29-NOV-2000) Institut fuer Immunogenetik,
Humboldt-Universitaet zu Berlin, Spandauer Damm 130, Berlin 14050,
Germany
FEATURES
source Location/Qualifiers
1..1365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.4"
/cell_type="peripheral blood leukocytes"
1..1365
/gene="ILT11"
121..1020
/gene="ILT11"
/note="Immunoglobulin superfamily receptor; similar to the
sequence in GenBank Accession Number AF212842"
/codon_start=1
/product="Immunoglobulin-like transcript 11 protein"
/protein_id="AAK52451.1"
/db_xref="GI:14028660"
/translation="MAPSHPSAQLQPGDVAVSPALMVLICGLSLGPRTHVQGNIL
SKATLMAEPGVSIRGNSVTIRCOGTLLEAGERYILVKEGSPFMDTONPLEPKNAFSP
IPSTHEHAGRGYCYYSAPAKSPSPDLELVNGFYNKPTLSALPSPVYVSCGNVTL
OCGRLELRDRILITLBECDHKLSKTLDSOLTSPSGPQALFPYGPVTPSHRMLRCYGRSR
LHILQWVSPSDLEILPYSGAADNLSPSQNSDSGTYASHLDYAVENLIRKGMAGLITLIL
VVLGILIFQDMHSORSPOAAAGR"
BASE COUNT 286 a 388 c 376 g 315 t
ORIGIN
alignment_scores:
Quality: 108.00 Length: 120

```

Ratio: 1.612 Gaps: 6
Percent similarity: 55.833 Percent Identity: 34.167

alignment_block:

US-09-471-276-831 x AF324830 ..

Align seg 1/1 to: AF324830 from: 1 to: 1365

```

1 MetSerMetLeuValValPheLeuLeuLeuThrPGLyValThrTrpGlyPr 17
   ::::: ::::: ::::: ::::: :::::
175 GTACAGCCCTCCCTCATGCTTCTGCTCGCTCGGCTGAGTCTGGCC 224
   ::::: ::::: ::::: ::::: :::::
17 oValThrGluAlaAlaIlePheTyrGluThrGln**SerLeuTrpAlaG 34
   ::::: ::::: ::::: ::::: :::::
225 CAGACCCAGCTGCAGGACGAGAACCTCCAAAGCCACCTCTGGGCTG 274
   ::::: ::::: ::::: ::::: :::::
34 LuSerGluHis**LeuLysThrLeuGlyGln..CysAspAlaAspValP 50
   ::::: ::::: ::::: ::::: :::::
275 AGCCAGGCTCTGTGATC...AGCCGGGGGAACCTGTGACCATTCGGTGT 321
   ::::: ::::: ::::: ::::: :::::
50 roGlyProProGlyAspSerArgLeuProAla..ValGlnGlu..... 63
   ::::: ::::: ::::: ::::: :::::
322 CAGCGACGACCTGGAGCCCGACGAATACCTGTGTTAAAGAGGAGACCC 371
   ::::: ::::: ::::: ::::: :::::
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIle 78
   ::::: ::::: ::::: ::::: :::::
372 AGAACCTCTGGACACACAGAACCCACTG.....GAGCCCAAGACA 412
   ::::: ::::: ::::: ::::: :::::
78 yHisIscInPheLeuLeuThrGlyAspThrGln.....GlyArgTyr 91
   ::::: ::::: ::::: ::::: :::::
413 AGCGCAGATTCCTCCATCCCATGATGCACAGACCATGACGAGAGATAC 462
   ::::: ::::: ::::: ::::: :::::
92 ArgCysArgSerGlyLeuSerThrGlyTrp**GlnLeuSerLysLeuLe 108
   ::::: ::::: ::::: ::::: :::::
463 CGCTGTACTACTACACAGCCCTGCAGGCTGTGTACAGCCGACGACCCCT 512
   ::::: ::::: ::::: ::::: :::::
108 uGluLeu 110
   ::::: ::::: ::::: ::::: :::::
513 GGAAGCTG 519

```

seq_name: gb_pr:HSU91925

```

seq_documentation_block:
LOCUS      HSU91925      1603 bp      mRNA      PRI      03-SEP-1997
DEFINITION Human clone HM18 monocytic inhibitory receptor precursor mRNA,
complete cds.
ACCESSION  U91925
VERSION    U91925.1
KEYWORDS   GI:2351798
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE  1 (bases 1 to 1603)
            Arm,J.P., Nwankwo,C. and Austen,K.F.
            Molecular identification of a novel family of human immunoglobulin
            superfamily members that possess immunoreceptor tyrosine-based
            inhibitory motifs and homology to the mouse gp9BI inhibitory
            receptor
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1603)
            Arm,J.P.
            Direct Submission
TITLE      Submitted (04-MAR-1997) Medline, Harvard Medical School and
JOURNAL    Brigham and Women's Hospital, 250 Longwood Avenue, Room 628,
            Boston, MA 02146, USA
FEATURES   Location/Qualifiers
            source
            1..1603
            /organism:"Homo sapiens"
            /db_xref:"taxon:9606"
            /chromosome:"19"
            /map:"19q13.4"
            /cell_type:"monocyte"

```

5'UTR
CDS
/clone:"HM18"
<1..45
46..1392
/codon_start:1
/product:"monocyte inhibitory receptor precursor"
/protein_id:"AAB6865.1"
/db_xref:"GI:2351799"
/translation:"MIPFTALLGLSLGPRTHMQAGPLPKPTLMAEBSVISMNS
VTWCGTLEAREYERLDRKESPAWDRONPLEPKKARFSIPMTEDYAGRYCYRS
PWGNSPDPLELYWTVGAYSKPTLSALSPVTSCKSVTLICQSPSPMDPTLLIKERA
AHPDLILSRSEHGAOOHQDAEPMSPTVSYGCTYKCFSSHGTSHTLLSHPSDFELIVS
GSLGPRSPRSVSTAAGPEODPLMPGVSFSGLRHMEVLLIGLVVSLILSLLL
ELLQIHWKQKHRTLAQOADPQRPGAEEPEPKDQKRSSPAADVQGENFCAAVK
NTPQEDGEMTDROSPHEDPOAVTYAVKHSRPREMASPPSPSGEFLTDKQAE
EDRQMDTEAASAPQDVTYQNLHSFTLRQKATERPPQEGASPAEPSVYATLAIH"
46..114
115..1389
/product:"monocyte inhibitory receptor"
115..822
/note:"encodes cytoplasmic domain"
823..885
/note:"encodes transmembrane domain"
886..1389
/note:"encodes cytoplasmic domain"
1117..1134
/note:"encodes ITIM motif"
1273..1290
/note:"encodes ITIM motif"
1393..1603
1578..1583
polyA_site
1603
/note:"14 A nucleotides"

BASE COUNT 384 a 517 c 427 g 275 t
ORIGIN

alignment_scores:
Quality: 103.50 Length: 121
Ratio: 1.617 Gaps: 7
Percent Similarity: 52.893 Percent Identity: 36.364

alignment_block:

US-09-471-276-831 x HSU91925 ..

Align seg 1/1 to: HSU91925 from: 1 to: 1603

```

2 SerMetLeuValValPhe.....LeuLeuLeuTrpGlyValThrTrpG 16
   ::::: ::::: ::::: ::::: :::::
43 GCCATGATCCCGACACCTTCAGGCTCTGCTCGCTCGGCTGAGTCTGG 92
   ::::: ::::: ::::: ::::: :::::
16 yProValThrGluAlaAlaIlePheTyrGluThrGln**SerLeuTrpA 33
   ::::: ::::: ::::: ::::: :::::
93 CCCCAGACCCACATGCGACGAGGCCCCCACCACCCACCTCTGGG 142
   ::::: ::::: ::::: ::::: :::::
33 LuSerGluHis**LeuLysThrLeuGlyGln..CysAspAlaAsp 48
   ::::: ::::: ::::: ::::: :::::
143 CTGAGCGACAGCTCTGTGATCAG...CTGGGGGAACCTGTGACCATCTG 188
   ::::: ::::: ::::: ::::: :::::
49 ValProGlyProProGlyAspSerArgLeuProAla..ValGlnGlu... 63
   ::::: ::::: ::::: ::::: :::::
189 GTGTACAGGGAGCCCTGGAGGCTCGGAGTACCGCTGATGAATAAGAGAAA 238
   ::::: ::::: ::::: ::::: :::::
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAla 76
   ::::: ::::: ::::: ::::: :::::
239 GCCCAGACCCCTGGACACAGACAGAACCCACTG.....GAGCCCAAG 279
   ::::: ::::: ::::: ::::: :::::
77 IleHisIscInPhe.....LeuLeuThrGlyAspThrGlnGlyAr 90
   ::::: ::::: ::::: ::::: :::::
280 AACAGAGCGAATTCCTCATCCATCCATGCACAGAGCATATGACGCGAG 329
   ::::: ::::: ::::: ::::: :::::
90 gTyrArgCysArgSerGlyLeuSerThrGlyTrp**GlnLeuSerLysL 107
   ::::: ::::: ::::: ::::: :::::
330 ATACCGCTGTACTATTCGACGCCCTGTGAGCTGTGCTGACGACCCGAGTAC 379

```

```

107 euleu1euleu 110
|||||
380 CCCTGGAGCTG 390

seq_name: gb_pat:AX046954

seq_documentation_block:
LOCUS AX046954 1625 bp DNA PAT 15-DEC-2000
DEFINITION Sequence 17 from Patent WO0068383.
ACCESSION AX046954
VERSION AX046954.1 GI:11876384
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1625)
Cosman,D.J., Anderson,D.M. and Borges,L.
Family of immunoregulators designated leukocyte immunoglobulin-like
receptors (LIR)
Patent: WO 0068383-A 17 16-NOV-2000;
JOURNAL IMMUNEX CORPORATION (US)
FEATURES
source location/Qualifiers
1..1625
/organism="Homo sapiens"
/db_xref="taxon:9606"
30..1376
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:11876385"
/protein_id="CAC18933.1"
/translation="MIPPTALICGLSGPRTMAGPLPKPTLMAEPGVSISMNS
VTMCOGTLAEKREYRLDKESPAEDRONPLEPKKARSLPSMTEDYAGRYCYRS
PVGWSQPSDPLELVTGAVSKPTLSALPSPLVTSKRSVTLDCOSRSPMDTFLIKERA
AHPILHLRSEHGADQHQAEFPMSPTSVHGCTYRCFSSHGESHYLLSHPSDLELIVS
GLSEGPSPPTRSVSTAAGPEODPLMPGSPVSHGLRRHMEVLICVLVYSILLSLIL
FLLOHMRGKHRTLAORADFORPPGAEPKRGGLORSSPADVOGEMFCAAVK
NTOPEDEVMTROSHPDEDOAVYAKVKHSRPREMASPSPISGEPFLDKQDAE
EDROMDTPEAASEAPODVTYARLHSFTLRKATEPPSPDEGASPAEPSTVATLAIH"
BASE COUNT 399 a 523 c 425 g 278 t
ORIGIN

alignment_scores:
Quality: 103.50 Length: 121
Ratio: 1.617 Gaps: 7
Percent Similarity: 52.893 Percent Identity: 36.364

alignment_block:
US-09-471-276-831 x AX046954 ..
Align seg 1/1 to: AX046954 from: 1 to: 1625
2 SerMetLeuValValPhe.....LeuLeuLeuTrpGlyValThrTrpG1 16
:::||||: ||||| ||||| ||||| |||||
27 GCCATGATCCCCACCTTCACGGCTCTGCTGCTCGGCTGAGTCTGGG 76
16 yProValThrGluAlaAlaIlePheTrpGlyThrGln**SerLeuTrpA 33
||||| ||||| ::||| |||||
77 CCCAGAGACCCACATGACAGGAGGCCCCCACCACCAACCCACCTCGGG 126
33 LacIuSerGluHis**LeuLysThrLeuGlyGln...CysAspAlaAsp 48
||||| ::||| ||||| |||||
127 CTGAGCGACGGCTCTGTGATCAG...CTGGGGGGAACCTCTGTACCATCTG 172
49 ValProGlyProProGlyAspSerArgLeuProAla..ValGluGlu... 63
||||| ||||| ||||| ||||| |||||
173 GTGTACAGGAGCCCTGGAGGCTCGGAGATACCGTCTGATTAAGAGGAAA 222
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAla 76
||||| ||||| ||||| ||||| |||||
223 GCCACGACCTTGGAGACAGACAGAACCCACTG.....GACCCCAAG 263

```

```

77 IleLysHisGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArg 90
||||| ::||| ||||| |||||
264 AACAGAGCCGATTCCTCATCCATGACAGACAGAGACTATTCAGAGGAG 313
90 gTyArgCysAr:SerGlyLeuSerThrGlyTrp**GlnLeuSerLysL 107
||||| ::||| ||||| |||||
314 ATACCGCTGTACTATTCGACGAGCCCTGTAGGCTGTGTACACCCAGTGACC 363
107 euleu1euleu 110
|||||
364 CCCTGGAGCTG 374

seq_name: gb_pr:AF025532

seq_documentation_block:
LOCUS AF025532 1625 bp mRNA PRI 01-DEC-1997
DEFINITION Homo sapiens leucocyte immunoglobulin-like receptor-5 (LIR-5) mRNA,
complete cds.
ACCESSION AF025532
VERSION AF025532.1 GI:2653870
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1625)
Borges,L., Hsu,M.-L., Fanger,N., Kubin,M. and Cosman,D.
A family of human lymphoid and myeloid immunoglobulin-like
receptors, some of which bind to MHC class I molecules
J. Immunol. (1997) In press
2 (bases 1 to 1625)
Borges,L., Hsu,M.-L., Fanger,N., Kubin,M. and Cosman,D.
Direct Submission
Submitted (18-SEP-1997) Molecular Biology, Immunex Corp., 51
University St., Seattle, WA 98101, USA
JOURNAL
FEATURES
source location/Qualifiers
1..1625
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1625
/gene="LIR-5"
30..1376
/gene="LIR-5"
/codon_start=1
/product="leucocyte immunoglobulin-like receptor-5"
/protein_id="AAB87666.1"
/db_xref="GI:2653871"
/translation="MIPPTALICGLSGPRTMAGPLPKPTLMAEPGVSISMNS
VTMCOGTLAEKREYRLDKESPAEDRONPLEPKKARSLPSMTEDYAGRYCYRS
PVGWSQPSDPLELVTGAVSKPTLSALPSPLVTSKRSVTLDCOSRSPMDTFLIKERA
AHPILHLRSEHGADQHQAEFPMSPTSVHGCTYRCFSSHGESHYLLSHPSDLELIVS
GLSEGPSPPTRSVSTAAGPEODPLMPGSPVSHGLRRHMEVLICVLVYSILLSLIL
FLLOHMRGKHRTLAORADFORPPGAEPKRGGLORSSPADVOGEMFCAAVK
NTOPEDEVMTROSHPDEDOAVYAKVKHSRPREMASPSPISGEPFLDKQDAE
EDROMDTPEAASEAPODVTYARLHSFTLRKATEPPSPDEGASPAEPSTVATLAIH"
BASE COUNT 399 a 523 c 425 g 278 t
ORIGIN

alignment_scores:
Quality: 103.50 Length: 121
Ratio: 1.617 Gaps: 7
Percent Similarity: 52.893 Percent Identity: 36.364

alignment_block:
US-09-471-276-831 x AF025532 ..
Align seg 1/1 to: AF025532 from: 1 to: 1625
2 SerMetLeuValValPhe.....LeuLeuLeuTrpGlyValThrTrpG1 16
:::||||: ||||| ||||| ||||| |||||
27 GCCATGATCCCCACCTTCACGGCTCTGCTGCTCGGCTGAGTCTGGG 76

```



```
REFERENCE 1 (bases 1 to 1818)
AUTHORS Colonna,M., Navarro,F., Bellon,T., Llano,M., Garcia,P.,
          Samaridis,J., Angman,L., Cella,M. and Lopez-Botet,M.
TITLE A common inhibitory receptor for major histocompatibility complex
        class I molecules on human lymphoid and myelomonocytic cells
JOURNAL J. Exp. Med. 186 (11), 1809-1818 (1997)
MEDLINE 98044246
REFERENCE 2 (bases 1 to 2018)
AUTHORS Colonna,M.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1997) Basel Institute for Immunology, 487
          Grenzachstrasse, Basel CH-4005, Switzerland
FEATURES
  source      Location/Qualifiers
              1..2018
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="19"
                /cell_type="myelomonocytic; NK; T; B"
                /map="19q13.4"
                1..1533
                  /note="ILT2c; alternatively spliced form of ILT2"
                  /codon_start=1
                  /product="Immunoglobulin-like transcript 2c"
                  /protein_id="AAC51881.1"
                  /db_xref="GI:2660706"
                  /translation="MPTLTVLICGLSLGSPRTYVQAGHLKPTLMAERGSVITQGS
                    VTLRQGGQETQEVRLYREKKTALMTTRIPQELVKKGPPISTWEHAGRICTYGS
                    DTAGSESSDPLELVNTGAYIKPTLSADPSVNVNGGVNIIQCDSQVAFEDSLKEG
                    EDHPQCLNSQPHARGSSRAIFSVGPVSPSRMWRCYAYDNSPYEWSLPDLELL
                    LVGSKKPSLSVQPGPIVAPBETLTJQCSGDAGYRNVLYKGEEDFLQLAGAQPOAG
                    LSOANFTLCGVSRYGGOYRCYGAHNLSEMSAPSDPDLIIAGOFYDRLSVQPGP
                    TVASGENVTLLCOSGMOTFLITKEGAADDPWRIRSTYOSQKYGAEPFGVTSNAHA
                    GTRYCGSSQSKPYLLTHRPDPLELVYSGPSGSPPTTGPSTSGPEDDPLTPTRSD
                    POSGCLRHAGVYIGILVAVALLLLELLLLFLILRHRRQKHWITSSPAQLPMPRKRTSM
                    LP"
BASE COUNT 428 a 665 c 549 g 376 t
ORIGIN
alignment_scores:
  Quality: 103.00 Length: 160
  Ratio: 1.288 Gaps: 9
Percent Similarity: 50.000 Percent Identity: 30.000
alignment_block:
US-09-471-276-831 x AF009007 ..
Align seg 1/1 to: AF009007 from: 1 to: 2018
1 MetSerMetLeuValValPheLeuLeuLeuTrrpGlyValThrtrpGlyPr 17
  |||::: ::::: ::::: |||::: |||::: ::::: |||:::
1 ATGACCCCTCCTCCTCAGCGTCGATCTGTCTGCTGAGTGGCCGCC 50
17 oValThrGluAlaAlaIlePheTyrGluThrGln**SerLeuTrrpAlaG 34
  | |||::: ::::: ::::: ::::: |||::: |||::: ::::: |||:::
51 CAGGACCCAGCTGTCAGGACGACGACCTCCCAACCCACCTCTGGCGCTG 100
34 luserGluHis**LeuLysThrLeuGly..GlnCysAspAlaAspValP 50
  |||::: ::::: |||::: |||::: |||::: |||::: |||:::
101 AACGAGCCTCTGTGATC...ACCCAGGGGAGGCTCTGTGACCTTCAGGTG 147
50 roGlyProProGlyAspSerArlgLeuProAlaValAl.GlnGlu..... 63
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
148 CAGGGGGGCCAGGACCCAGAGTACCGCTCATATATAGAGAAAGAAAC 197
64 .....TrrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
  |||::: ::::: |||::: |||::: |||::: |||::: |||:::
198 AGCAGCTGTGATTAACGATCCACAGAGACTT.....GTGAGA 238
78 yshIsGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArgTyr 91
  || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 AGGGCCAGTTCCCATCCATCCATCAGCTGGAGAACATGACGGCGGTAT 288
```

```
92 ArgCysArgSerGlyLeuSerThr...GlyTrrp**GlnLeuSerLysLe 107
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
289 CCCTGTACTACTATGTTAGCGACACTGCGACGCCGCTCAGAGACTGACCC 338
107 uLeuGlnLeuThr.....GlyProLysValIleLeuAlaCys... 118
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339 CCTGAGCTGTGGTGTGACAGGAGCCTACATCAACCCACCTCTCAGCCC 388
112 .....GlyProLysValIleLeuAlaCys... 118
389 AGCCACACCCCGCTGTCGACTCAGGAGGAAATGATTCCTCAGCTTGAC 438
119 ...SerLeuAlaLeuAspGlyAlaSer 126
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
439 TCACAGCTGCCATTTCATGCTTCAGT 465
seq_name: gb_pr:AF009005
seq_documentation_block: 2094 bp mRNA PRI 03-DEC-1997
LOCUS AF009005
DEFINITION Homo sapiens Immunoglobulin-like transcript 2a mRNA, complete cds.
ACCESSION AF009005
VERSION AF009005.1 GI:2660701
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2094)
AUTHORS Colonna,M., Navarro,F., Bellon,T., Llano,M., Garcia,P.,
          Samaridis,J., Angman,L., Cella,M. and Lopez-Botet,M.
TITLE A common inhibitory receptor for major histocompatibility complex
        class I molecules on human lymphoid and myelomonocytic cells
JOURNAL J. Exp. Med. 186 (11), 1809-1818 (1997)
MEDLINE 98044246
REFERENCE 2 (bases 1 to 2094)
AUTHORS Colonna,M.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1997) Basel Institute for Immunology, 487
          Grenzachstrasse, Basel CH-4005, Switzerland
FEATURES
  source      Location/Qualifiers
              1..2094
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="19"
                /cell_type="myelomonocytic; NK; T; B"
                /map="19q13.4"
                1..1953
                  /note="ILT2a; allelic form of ILT2"
                  /codon_start=1
                  /product="Immunoglobulin-like transcript 2a"
                  /protein_id="AAC51879.1"
                  /db_xref="GI:2660702"
                  /translation="MPTLTVLICGLSLGSPRTYVQAGHLKPTLMAERGSVITQGS
                    VTLRQGGQETQEVRLYREKKTALMTTRIPQELVKKGPPISTWEHAGRICTYGS
                    DTAGSESSDPLELVNTGAYIKPTLSADPSVNVNGGVNIIQCDSQVAFEDSLKEG
                    EDHPQCLNSQPHARGSSRAIFSVGPVSPSRMWRCYAYDNSPYEWSLPDLELL
                    LVGSKKPSLSVQPGPIVAPBETLTJQCSGDAGYRNVLYKGEEDFLQLAGAQPOAG
                    LSOANFTLCGVSRYGGOYRCYGAHNLSEMSAPSDPDLIIAGOFYDRLSVQPGP
                    TVASGENVTLLCOSGMOTFLITKEGAADDPWRIRSTYOSQKYGAEPFGVTSNAHA
                    GTRYCGSSQSKPYLLTHRPDPLELVYSGPSGSPPTTGPSTSGPEDDPLTPTRSD
                    POSGCLRHAGVYIGILVAVALLLLELLLLFLILRHRRQKHWITSSPAQLPMPRKRTSM
                    LP"
BASE COUNT 448 a 686 c 574 g 386 t
ORIGIN
alignment_scores:
  Quality: 103.00 Length: 160
  Ratio: 1.288 Gaps: 9
Percent Similarity: 50.000 Percent Identity: 30.000
```

alignment_block:

US-09-471-276-831 x AF009005 ..

Align seg 1/1 to: AF009005 from: 1 to: 2094

```

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyPr
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1 ATGACCCCCCTCCTCAGCGCTCATCTCTCTCGCGCTGAGTGGCCCC 50
17 oValThrGluAlaAlaIlePheTrpGluThrGln**SerLeuTrpAlaG 34
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
51 CAGCAGCCAGCTCAGGACGACGACCTCCCAAGCCCACTCTGGGCTG 100
34 IuSerGluHis**LeuLysThrLeuGly..GlnCysAspAlaAspAlp 50
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
101 AACGAGCGCTCTGTGATC...ACCGAGGGAGCTCTGTGACCCCTCAGGT 147
50 roGlyProProGlyAspSerArgLeuProAlaVal.GlnGlu..... 63
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
148 CAGGGGGCCAGAGACCCAGAGTACCGCTCTATATAGAGAAAGAAAC 197
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
198 AGCACTCTGGATTACACGATCCACAGAGCTT.....GTGAGA 238
78 yshIscGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArgTyr 91
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
239 AGGGCCAGCTTCCCATCCATCCATCCATCCGGAACATGACAGCGCGGTAT 288
92 ArgCysArgSerGlyLeuSerThr..GlyTrp**GlnLeuSerLysLe 107
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
289 CGCTGTACTATGTGTAGCAGACACTGACGCCCTCAGAGAGAGTGACCC 338
107 uLeuGluLeuThr.....GlnProLysValLeuAlaCys... 118
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
339 CCTGAGCTGTGTGTGACAGAGACCTACATCAAAACCCCTCTCAGCCC 388
112 .....GlyProLysValLeuAlaCys... 118
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
389 AGCCAGCCCGCCGTGTGAACCTCAGAGGGAATGTAATCTCTCAGGTGAC 438
119 ...SerLeuAlaLeuAspGlyAlaSer 126
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
439 TCACAGGTGGCATTTGATGCTTCACGT 465
seq_name: gb_pat:AX046940
seq_documentation_block:
LOCUS AX046940 2777 bp DNA PAT 15-DEC-2000
DEFINITION Sequence 3 from Patent WO0068383.
ACCESSION AX046940
VERSION AX046940.1 GI:11876370
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 2777)
Cosman,D.J., Anderson,D.M. and Borges,L.
Family of immunoregulators designated leukocyte immunoglobulin-like
receptors (IL1r)
Patent: WO 0068383-A 3 16-NOV-2000;
JOURNAL IMMUNEX CORPORATION (US)
FEATURES
Location/Qualifiers
source
1..2777
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
168..2126
/note:"unamed protein product"
/codon_start:1
/protein_id:"CAC18927.1"
/db_xref:"GI:11876371"

```

```

/translation:"MPTLTVLJCLGLSGPPTHVQAGHLPEKPLMAEPGSVITGSP
VTRLCGGGEOEYELRYREKKTALITRIPDLVYKGGFPIPSIMHEAGRYCYGS
DTAGSSSDPLELVPGAYIKPTLSAPSPVNVGSGGVLIQCDQVAPPDGSLCKBG
EDERQCLNSOPHARGSSKALFSVGPVSPSRMMWRKCAIYDSNSTEYSLSDLELL
VLGVRKPSLSVQDP IVAPEETLTGCGSDGVRVLYIKDGERDPLQLAGAPQAG
LSQANFTLGPVSRSYSGQYRCYGAINHLSSEMSAPSDPLDI LAGQFYDRVSLSVQPG
TVASGNVTLTLCOSQMMQTFLLTEGAADPWRRLRSTYQSKYQAEPMGVTSAHA
GYRCYGSQSSKPYLLTHPSDPLELVGSPSCGSPSPPTGPKSTSGAPEDPLPTGS
DPSGLGHILGVIC ILVAAILLILLLILFLILBRROGKHMSTORAKADQHPHGA
VCPRETDGRLKRRSPBADAQENIYAAVKTQDPDGYEMDTROSPHEDDQAVTYAE
VKHSRPRKEMASPSPLSGEFLTKDRQAEEDRMDTEAASEARQDVTYVQLSLTL
RREATERPPSGEPPAVPSIYATLAIH"
BASE COUNT 657 a 840 c 754 g 526 t
ORIGIN

```

```

alignment_scores:
Quality: 103.00 Length: 160
Ratio: 1.288 Gaps: 9
Percent Similarity: 50.000 Percent Identity: 30.000

```

alignment_block:

US-09-471-276-831 x AX046940 ..

Align seg 1/1 to: AX046940 from: 1 to: 2777

```

1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
168 ATGACCCCCCTCCTCAGCGCTCATCTCTCTCGCGCTGAGTGGCCCC 217
17 oValThrGluAlaAlaIlePheTrpGluThrGln**SerLeuTrpAlaG 34
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
218 CAGCAGCCAGCTCAGGACGACGACCTCCCAAGCCCACTCTGGGCTG 267
34 IuSerGluHis**LeuLysThrLeuGly..GlnCysAspAlaAspAlp 50
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
268 AACGAGCGCTCTGTGATC...ACCGAGGGAGCTCTGTGACCCCTCAGGT 314
50 roGlyProProGlyAspSerArgLeuProAlaVal.GlnGlu..... 63
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
315 CAGGGGGCCAGAGACCCAGAGTACCGCTCTATATAGAGAAAGAAAC 364
64 .....TrpGlyAlaGlnGluProValHisLeuAspSerProAlaIleL 78
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
365 AGCACTCTGGATTACACGATCCACAGAGCTT.....GTGAGA 405
78 yshIscGlnPhe.....LeuLeuThrGlyAspThrGlnGlyArgTyr 91
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
406 AGGGCCAGCTTCCCATCCATCCATCCATCCGGAACATGACAGCGCGGTAT 455
92 ArgCysArgSerGlyLeuSerThr..GlyTrp**GlnLeuSerLysLe 107
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
456 CGCTGTACTATGTGTAGCAGACACTGACGCCCTCAGAGAGAGTGACCC 505
107 uLeuGluLeuThr.....GlnProLysValLeuAlaCys... 118
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
506 CCGGAGCTGTGTGTGACAGAGACCTACATCAAAACCCCTCTCAGCCC 555
112 .....GlyProLysValLeuAlaCys... 118
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
556 AGCCAGCCCGCCGTGTGAACCTCAGAGGGAATGTAATCTCTCAGGTGAC 605
119 ...SerLeuAlaLeuAspGlyAlaSer 126
  |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
606 TCACAGGTGGCATTTGATGCTTCACGT 632
seq_name: gb_pr:AF004230
seq_documentation_block:
LOCUS AF004230 2897 bp mRNA PRI 27-AUG-1997
DEFINITION Homo sapiens monocyte/macrophage Ig-related receptor MIR-7 (MIR
cl-7) mRNA, complete cds.
ACCESSION AF004230

```



```

4310 .....GAGCCCAAGAACAGCCAGATTCATCCATCATGACAGAG 4354
86 AspThrGlnGlyArgTyrArgCysArgSerGlyLeuSerThrGlyTrp** 102
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4355 GACTATGACGAGGAGATACCGCTGTACTATCCAGCCCTGTAGGCTGTC 4404
102 *GlnLeuSerLysLeuLeuGlnLeuThr..... 111
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4405 ACAGCCCACTAGACCCCTGGAGCTGGTATGACAGGTGAGAGACACTCA 4454
112 ..GlyProLysValLeuAlaCysSerLeu.....AlaLeuAspGly 124
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4455 GGGGTCCAGACCCCGAGCTCTGCCCTCAGAGAGGGGTGCTCTCAGGG 4504
125 AlaSer 126
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4505 GCATCT 4510

seq_name: gb_pr:AC011515

seq_documentation_block:
LOCUS AC011515 47084 bp DNA PRI 29-JUN-2000
DEFINITION Homo sapiens chromosome 19 clone LLNLF-139A9, complete sequence.
ACCESSION AC011515
VERSION AC011515.4 GI:8810258
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 47084)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 47084)
DOE Joint Genome Institute.
AUTHORS Direct Submission
JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 47084)
DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE
AUTHORS Direct Submission
JOURNAL Submitted (29-JUN-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jun 29, 2000 this sequence version replaced gi:8576078.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence:
Estimated Total Number of Errors is 0.
FEATURES
source
1. 47084
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="LLNLF-139A9"
BASE COUNT 12882 a 10887 c 10717 g 12598 t
ORIGIN
alignment_scores:
Quality: 103.00 Length: 136
Ratio: 1.537 Gaps: 8
Percent Similarity: 49.265 Percent Identity: 33.824
alignment_block:
US-09-471-276-831 x AC011515 ..
Align seg 1/1 to: AC011515 from: 1 to: 47084
12 G|ValThrTrpGlyProValThrGluAlaAlaIlePheTyrGluThrG1 28
||| ::||| |||:::

```

```

7484 GGAAGGCTTGAGGATCCAGCCTCGATTTCTTCAGAGCCCTCCCCAA 7533
28 n***SerLeuTrpAlaGluSerGlnHis***LeuLysThrLeuGlyGln 44
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
7534 ACCACCTCTGGCTGGAGGACGAGCCAGCTCTGTATCAG...CTGGGGGAC 7579
45 ..CysAlaHisAspValProGlyProProGlyAspSerArgLeuProAla 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7580 TCTGTGACACTCTGCTGTCTCAGGGACCTTGAGGCTGGAGATACCTCT 7629
61 ValGlnGlu.....TrpGluAlaGlnGluProValHisL 72
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
7630 GGATTAAGAGAAACCCACACCTGAGACAGACAGAACACCACTG... 7675
72 euAspSerProAlaIleLysHisGlnPhe.....LeuLeuThrGly 85
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
7676 .....GAGCCCAAGAACAGCCAGATTCATCCATCATGACAGAG 7720
86 AspThrGlnGlyArgTyrArgCysArgSerGlyLeuSerThrGlyTrp** 102
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7721 GACTATGACGAGGAGATACCGCTGTACTATCCAGCCCTGTAGGCTGTC 7770
102 *GlnLeuSerLysLeuLeuGlnLeuThr..... 111
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7771 ACAGCCCACTAGACCCCTGGAGCTGGTATGACAGGTGAGAGACACTCA 7820
112 ..GlyProLysValLeuAlaCysSerLeu.....AlaLeuAspGly 124
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7821 GGGGTCCAGACCCCGAGCTCTGCCCTCAGAGAGGGGTGCTCTCAGGG 7870
125 AlaSer 126
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7871 GCATCT 7876

seq_name: gb_hlg:AL139117

seq_documentation_block:
LOCUS AL139117 148497 bp DNA HTG 19-APR-2001
DEFINITION Homo sapiens chromosome 19 clone RPL-52N12, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL139117
VERSION AL139117.8 GI:12189412
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 148497)
DOE Joint Genome Institute and Sanger Centre, Hinxton, Cambridgeshire,
UK.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jan 13, 2001 this sequence version replaced gi:9796322.
COMMENT ***** Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
***** Project Information
Center project name: dj52N12
***** Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator ABI; 61% of reads
Chemistry: Dye-terminator Big Dye; 35% of reads
Chemistry: Dye-primer-amerasham; 2% of reads
Chemistry: Dye-primer-Big Dye; 0% of reads
Consensus quality: 147416 bases at least Q40
Consensus quality: 148180 bases at least Q30
Consensus quality: 148423 bases at least Q20
Insert size: 148497; sum-of-contents
Insert size: 146392; agarose-fp

```



```

57 ArgLeuProAla.ValGlnGlu.....TrpGlyAlaGln 68
   ::::: :::::
168 GGAAGTACCGCTGGATTAAGAGAAAGCCAGCACCCTGGGACAGACAGA 217
   ::::: :::::
68 LeuProValHisLeuAspSerProAlaIleLysHisGlnPhe..... 81
   ::::: :::::
218 ACCCACTG.....GAGCCCAAGAACAGGACGATTCCTCATCCCA 258
   ::::: :::::
82 LeuLeuThrGlyAspThrGlnGlyArgGlyArgCysArgSerGlyLeu 98
   ::::: :::::
259 TCCATACACAGAGCATATCCAGCAGATACCGCTGTTACTATCCAGCCC 308
   ::::: :::::
98 rThrGlyTrp**GlnLeuSerLysLeuLeuGluLeu 110
   ::::: :::::
309 TGTAAGCTGTGTCACAGCCAGTACCCCTGGAGCTG 345
seq_name: gb_pr:AC008984

```

seq_documentation_block:

```

LOCUS AC008984 39829 bp DNA PRI 18-APR-2000
DEFINITION Homo sapiens chromosome 19 clone LLNF-191E4, complete sequence.
ACCESSION AC008984
VERSION AC008984.5 GI:6984369
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 39829)
DOE Joint Genome Institute.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 39829)
DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 39829)
DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Feb 17, 2000 this sequence version replaced gi:6910557.
COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated Total Number of Errors is 0.1.

```

FEATURES

```

source
1..39829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="LLNF-191E4"
BASE COUNT 10523 a 9405 c 9541 g 10360 t
ORIGIN

```

alignment_scores:

```

Quality: 102.00 Length: 117
Ratio: 1.569 Gaps: 7
Percent Similarity: 55.556 Percent Identity: 34.188

```

alignment_block:

```

US-09-471-276-831 x AC008984

```

Align seg 1/1 to: AC008984 from: 1 to: 39829

```

12 GlyValThrTrpGlyProValThrGluAlaIlePheThrGlu 28
   ::::: :::::
21625 GGGAGGCTTTGGAGCCAGACTGTGATTTCTTCAGGAACTCTCCAA 21674
   ::::: :::::
28 n***SerLeuTrpAlaGluSerGluHis**LeuLysThrLeuGlyGln 44
   ::::: :::::
21675 AGCCACCTCTGGGCTGAGCCAGGCTCTGTGATC...AGCGGGGAGACT 21721
   ::::: :::::
45 CysAspAlaAspValProGlyProProGlyAspSerArgLeuProAla 60
   ::::: :::::
21722 CTGTGACCATTCGCTGTGACGAGCCTGGAGCCCAAGATACCGTCTG 21771
   ::::: :::::
61 ValGlnGlu.....TrpGlyAlaGlnGluProValHisLeu 72
   ::::: :::::
21772 GTTAAAGAGGAGAACCCAGAACCTGGGACACAGACGAACTGCTG... 21816
   ::::: :::::
72 uAspSerProAlaIleLysHisGlnPheLeuThrGlyAspThrGln 88
   ::::: :::::
21817 ...GAGCCCAAGAACAGCCAGATTCCTCATCCATCCATGACAGAC 21862
   ::::: :::::
89 .....GlyArgTyrArgCysArgSerGlyLeuSerThrGlyTrp** 102
   ::::: :::::
21863 ACATGACAGGAGATACCGCTGTTACTACTACAGCCCTGACGCTGGTCA 21912
   ::::: :::::
103 GlnLeuSerLysLeuGluLeu.....ThrGlyProLysValLeu 116
   ::::: :::::
21913 GAGCCAGCCAGCCCTGAGCTGTGTGTCACAGCTGAGAGACACTC 21960
   ::::: :::::
seq_name: gb_htg:AL355172

```

seq_documentation_block:

```

LOCUS AL355172 97137 bp DNA HTG 21-APR-2001
DEFINITION Homo sapiens chromosome 19 clone RP3-598H20, *** SEQUENCING IN
PROGRESS *** 6 unordered pieces.
ACCESSION AL355172
VERSION AL355172.6 GI:13785046
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Milne,S.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonequest@sanger.ac.uk
On Apr 24, 2001 this sequence version replaced gi:13624978.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- project information
Center project name: dJ598H20
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Chemistry: Dye-primer Big Dye; 0% of reads
Consensus quality: 92366 bases at least Q40
Consensus quality: 94380 bases at least Q30
Consensus quality: 95853 bases at least Q20
Insert size: 96637; sum-of-contigs
Insert size: 96776; 3.8% error; agarose-fp
Quality coverage: 6.86x in Q20 bases; sum-of-contigs quality
coverage: 8.38x in Q20 bases; agarose-fp
-----

```

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is


```

66 aglnGlu.....ProValHisLeuAsp.... 73
|||||
26190 CCAGGACTACCACTGGATTAAGAGGAGCCAGCCCTTGACAGAA 26239
74 .....SerProAlaIleIysHisGlnPheLeuLeuThrGlyAsp 86
|||||
26240 ATAAACCCAGTCGAAACCAAGACAGATTCTCCATCCCATCCATG 26289
7ThcGln.....GlyArgTyrArgCysArgSerGlyLeuSerThrG1 100
|||||
26290 ACACAGACACCATGACGAGAGTACCGCTGCCACATATTACAGCTCTGCAGG 26339
100 YTPP**GlnLeuSerLysLeuLeuGlnLeuThr..... 111
|||||
26340 CTGCTCAGAGCCGACGCCCTCGAGCTGTGTATGACAGCTGAGAGA 26389
112 .....GlyProLysValLeuAlaCysSerLeu.....AlaLeu 122
|||||
26390 CACTCTGGGCTCCGACCCAGCCGCTGCTCCAGAAAGCGGTCCGCTC 26439
123 AspGlyAlaSer 126
26440 TCAGGCGCGTCT 26451
seq_name: gb_pr:AF041262

seq_documentation_block:
LOCUS AF041262 1395 bp mRNA PRI 05-JAN-1999
DEFINITION Homo sapiens Immunoglobulin-like transcript 8 mRNA, complete cds.
ACCESSION AF041262
VERSION AF041262.1 GI:4104892
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
COLONNA,M.
TITLE Immunoglobulin-like transcript 8
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1395)
REFERENCE 2 (bases 1 to 1395)
JOURNAL Submitted (07-JAN-1998) Basel Institute for Immunology, 487
TITLE Direct Submission
JOURNAL Grenzacherstrasse, Basel CH-4005, Switzerland
FEATURES
source
1..1395
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome:"19"
/map:"19q13.4"
/cell_type="Leukocyte"
/clone:"3"
1..1395
/function="activating receptor"
/note="Immunoglobulin superfamily member; ILT8"
/codon_start:1
/product="Immunoglobulin-like transcript 8"
/protein_id="AAD02204.1"
/db_xref="GI:4104893"
/transcript_id="M7PILTVL1CIGLSIGPHTVQACPPFKPTLMAEPGSAVSWGSP
VITWCGSLAEAEYRLDKESPEMDRNPNLEPKKAFESIISTEHHIAGRYCHAYS
SAGSESPDALEIVMTGAYSKPTLSALPSPVVASGGMPLQCGSGKGYHOPVLMKEGE
HOLPRTLSQOLHSGCPALPVPVGNVSHRRRTCYIYNNKTPPVNHSHPDLEILP
SGVSRKPSLTLQGFVLAPOSULTQCGSDVGYDFVLYKDEERFLQRPQOPQAGL
SQANFTLIPVSPSHIGGOYRCYGAHNLSEWSAPSDPLILNAGQIYDVPSVSAQPGPT
VASGCVNVLICOSMMQFDTFLITKEGAHAPRLRLSGAHKGYQAEFPMSVYTSAHAG
TYRCGYSYSSNPHLSPSPSEPLELWVSASHAKDYVENLIRMGAGLVIVFLGILFLE
AHSORNPQDAARR"
BASE COUNT 292 a 463 c 381 g 259 t
ORIGIN

```

```

alignment_scores:
  Quality: 94.50      Length: 126
  Ratio: 1.524      Gaps: 5
  Percent Similarity: 49.206      Percent Identity: 28.571
alignment_block:
US-09-471-276-831 x AF041262 ..
Align seg 1/1 to: AF041262 from: 1 to: 1395
1 MetSerMetLeuValValPheLeuLeuLeuTrpGlyValThrTrpGlyTr 17
|||||
1 ATGACCCCATCCTCCACGCTGATCTCTCGCGCTGAGTCTGCGCCC 50
17 ovalThrcGluAlaIlePheThrGluThrGln**SerLeuTrpAlaG 34
| |||||
51 CAGACCCACAGCTGAGCAGAGGCCCTTCCCAACCCACCCCTCTGGGCTG 100
34 Lu.....SerGlnHis**LeuLysThrLeu 42
||
101 AGCCAGGCTCTGTGATCAGCTGGGGAGCCCGTGACCATCTGCTGACG 150
43 GlyGlnCysAspAla.....AspValProGlyProProG1 54
|||||
151 GCGAGCCTGAGGCCCGCAGAGTACCGACCTGATTAAGAGGAGCCCA.. 198
54 YASPserArgLeuProAlaValGlnGluTrpGlyAluGlnGluProValH 71
|||||
199 .....GAGCCCTGGACAGAAATTAACCCACTG.. 225
71 IsLeuAspSerProAlaIleIysHisGlnPheLeuLeuThrGlyAspThr 87
|||||
226 .....GAACCCAGAACAGAGCCGAGATTCTCCATCCCATCATACAA 267
88 Gln.....GlyArgTyrArgCysArgSerGlyLeuSerThrGlyTr 101
|||||
268 GAGCACCATCGCGGAGATVACCCCTGCCACTATTACAGCTGTCAGAGCTG 317
101 P***GlnLeuSerLysLeuLeuGlnLeu 110
|||||
318 GTCAGAGCCGACGACGCCCTGAGCTG 345
seq_name: gb_pr:AF322217

seq_documentation_block:
LOCUS AF322217 1745 bp mRNA ROD 27-APR-2001
DEFINITION Rattus norvegicus Inhibin binding protein short isoform mRNA,
complete cds; alternatively spliced.
ACCESSION AF322217
VERSION AF322217.1 GI:13810899
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
TITLE 1 (bases 1 to 1745)
REFERENCE 1 (bases 1 to 1745)
AUTHORS Bernard,D.J. and Woodruff,T.K.
JOURNAL Inhibin Binding Protein In Rats: Alternative Transcripts and
Mol. Endocrinol. 15 (4), 654-667 (2001)
JOURNAL 11266515
PUBMED 2 (bases 1 to 1745)
REFERENCE 2 (bases 1 to 1745)
AUTHORS Bernard,D.J. and Woodruff,T.K.
JOURNAL Direct Submission
JOURNAL Submitted (18-NOV-2000) Neurobiology and Physiology, Northwestern
University, 2153 North Campus Drive, Evanston, IL 60208, USA
FEATURES
source
1..1745
location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
170..871
CDS

```



```

/note="InhibP.S; alternatively spliced"
/codon_start=1
/product="inhibin binding protein short isoform"
/protein_id="AKK40084.1"
/db_xref="GI:13810900"
/translation="MMLRTFTLLCLMNLNRCMTMAAVESPELMIIESNPQAPWEN
ITLMKSPSRVSKFLLKLDNTQMTWHPPTFPVSGALTSNGLYSCYWN
KGMKSPKILEEAPGLPKPIEMFOAETPPGCVNVLICGMLQDLVFMLEKGYT
EPIDYOVPTGMAIRSIDNLAENSGIYICRTHIMLPINSEPSNPLKLVYAGCAH
GCMHLTVIPIGIMAG"
BASE COUNT      481 a      436 c      345 g      483 t
ORIGIN
101 rp***GlnLeuSerLysLeuGlnLeuThrGlyPro 113
483 GGTGGAAGCCCAAGTAATCTAGAGCTAGAGACCA 520
seq_name: gb_pr:AF031554
seq_documentation_block:
LOCUS      AF031554      2066 bp      mRNA      PRI      09-DEC-1997
DEFINITION Homo sapiens clone 17.18 Immunoglobulin-like transcript 5 mRNA,
complete cds.
ACCESSION  AF031554
VERSION     AF031554.1  GI:2665642
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 2066)
AUTHORS     Colonna,M., Navarro,F., Bellon,T., Llano,M., Garcia,P.,
Samiridis,J., Angman,L., Cella,M. and Lopez-Botet,M.
TITLE       A common inhibitory receptor for major histocompatibility complex

```

```

JOURNAL      J. Exp. Med. 186 (11), 1809-1818 (1997)
MEDLINE      98044246
REFERENCE    2 (bases 1 to 2066)
AUTHORS     Colonna,M.
TITLE       Direct Submision
JOURNAL      Submitted (27-OCT-1997) Basel Institute for Immunology, 487
Grenzacherstrasse, Basel 4005, Switzerland
FEATURES
source
1..2066
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.4"
/cell_type="myelomonocyte"
/clone="17.18"
1..1901
/note="Inhibitory receptor; ILT5"
/codon_start=3
/product="Immunoglobulin-like transcript 5"
/protein_id="AAC51901.1"
/db_xref="GI:2665643"
/translation="MTPALVALLCGLSLGPRTRYQAGPPRTMAEPGVISWGSPP
VTIMCGSLEAOEYQLDKESPEPMDRNRPDLPRKNARFISMTQHNAGRYCHYS
SAGWSESPDPELVMTGAYSKPTLSLPSVVAAGSGNMTLRGSKRYHNHFLMEKE
HOLPTDSDQIHSGFQALPPVGVNPSHRARPTCYUYNMTPOVMSPLEILP
SVSRKRPILLTIGQRYLAPGSLTLQCGSDVGYDRPLYKEBERFLOPGQPPACL
SOANFTLGPVSPSNGQYRCYGAHNISSEMSAPSDLNTLMAGQITDYVISAPQRP
VASGENVTLLCSWQFDFTLTKEDANRPLRLSRMGANRYQAEFWSPTSAHAG
TYRCGYSSNBNHLISHPSEPLELWISGSGSLPPGPPSTPGLRILEVLIGVS
AFVLLFLFLRLRORHSKRTSQRKTDORPAGAETPRDGLLRSSPADV
OENLYAANKRKDQSEDRVELDSQSDHDEPQAVYAPVKKHSPPREMASSPSISGE
FLDTKDRQYEDRQMDTEAASDSDVYTAQLHSITLIRKATTEPPSGEPPADPS
ITATLAIR"
BASE COUNT      451 a      691 c      565 g      359 t
ORIGIN
101 MetSerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
3 ATGACGCCCGCCCTCACAGCCCTGCTTGGCTTGGCTGAGTCTGGGCC 52
17 ova1ThrGlnAlaAlaIlePheTyrgluThrGln***SerLeuTrpAla 34
53 CAGACCGCGCTGCAGCAGGAGCCCTCCCAACCAACCCCTGCGGCTG 102
34 lu.....SerGluHis***LeuLysThrLeu 42
103 AGCCAGGCTGTGTATCAGCTGGGAGCCCGTGACCATCTGTGTGCAG 152
43 GlyGlnCysAspAla.....AspValProGlyProGly 54
153 GCGAGCCTGGAGGCCAGAGTACCACCTGATTAAGAGGGAAGCCCA.. 200
54 yAspSerArgLeuProAlaValGlnGluTrpGlyAlaGlnGluProVal 71
201 .....GAGCCCTGGACAGAAATAAACCCACAGC. 227
71 isLeuAspSerProAlaIleLysHisGlnPheLeuLeuThrGlyAspThr 87
228 .....GAACCCAGAAAGAGCCAGATTCCTCCATCCATCATGACA 269
88 Gln.....GlyArgTyrglyCysArgSerGlyLeuSerThrGlyTr 101

```

```

111          |||
270 CAGCACCATGCAGGAGATACCCCTGACATTTACAGCTCTGCAGCGTG 319
101 P***GlnLeuSerLysLeuLeuGluLeu 110
|||||:|||||
320 GTCAGAGCCAGCAGCCCTGCAGCGTG 347

```

seq_name: gb_fr:AF322216

seq_documentation_block:

LOCUS AF322216 4407 bp mRNA ROD 27-APR-2001
DEFINITION Rattus norvegicus Inhibin binding protein long isoform mRNA,
complete cds; alternatively spliced.

ACCESSION AF322216
VERSION AF322216.1 GI:13810897

KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 4407)

AUTHORS Bernard,D.J. and Woodruff,T.K.
TITLE Inhibin Binding Protein in Rats: Alternative Transcripts and
Regulation in the Pituitary across the Estrous Cycle
JOURNAL Mol. Endocrinol. 15 (4), 654-667 (2001)

MEDLINE 21168090
PubMed 11266515

REFERENCE 2 (bases 1 to 4407)
AUTHORS Bernard,D.J. and Woodruff,T.K.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-2000) Neurobiology and Physiology, Northwestern
University, 2153 North Campus Drive, Evanston, IL 60208, USA

FEATURES
source location/Qualifiers

```

CDS
1..4407
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
172..4134
/note="Inhibin-L: alternatively spliced"
/codon_start=1
/product="Inhibin binding protein long isoform"
/protein_id="AAK40083.1"
/db_xref="GI:13810898"
/translation="MMLRTFTLLLCIMLNRCMTSMAAVESQPELIMESNPQAPMEN
ITLMCKSPRSVSKFLKDKNTOMTHPPYTFQVSPFTGALTESNTGLYRCQWNE
KGMKSPKLELEARGOLPKPIEMIOAEPPRCNVNITICGMLODYFMFKCYT
EPIDYQVPTGTAIFSIDNLAPENSGITICRTHOMLPKMEPSNPPLKLVAGLTPK
PULTAHPCPTLAPGESLSLRCSGLDILKIMVDTPKTMLVQPSPIQMGONVSLKC
LKTQDGHVLCFYDGSYRGSLDILKIMVDTPKTMLVQPSPIQMGONVSLKC
GGLMGVGLAHKKGEEKPLQFLDATSNTGNSFPLKNVTLVEFLEEMETTFOKFSVDG
IKMATYNTVELLVAMPSSVFEVKTITLQCVSHPLVEFLEEMETTFOKFSVDG
FITNIEGGTGTSCSYRIESHANIMRSRPLKLVPAVNGPLPMSILDEAIRVS
LTVQPLSLLLVLMQHKRCRLRLREANILGTAQVAMLYIITALLCGLCGALTEB
LEITMPTPELMAETNFPQAFKNTVLMCRSPSGSTFEFVLKDGTEWIAVRPASEQ
VRAAPLALQSHSGSYHCHSEMAVSESEALELVGTDLPRPVISAPLPICQEL
LOIRCKGLEGEGFALYKMGDEPQQQLGAVREAFPIQRMEDKEENYSCRTKOE
QPFKSEPSAPLELVKELPEKPFKWTASPVYTCRVTENCSTREHMSPIILYKDG
NEIASDPMWGNPSTAHPLISVIGIDGYSORYDYSDSTWSPSPDVELIIEFPE
PKPTLAPGPPVLPCKNVLKCGCIPOGMRRALLQEGAHAPLOQASAGTSVDLLH
TVGADFGNISCYVYETTMNSNGSYLSTPLMIVWDTFPEWLSPESSVVMGQNT
LMCGPVMHGVGVIHAKEGEATSMQLMDSTSNQAPPIINISGASIGRSCCHPDMMS
PIKIQPSNTLELVGLLKPSPILVQPGPMVPGNMFPCCGGLPDSFVYLKCGTO
QPIEQQRSGYRADFMVPRDQSGVSCYVYLASAPLVASNHNSNLEIWTADPK
PSLMSPTVFEKIDITLQCRGPLGVEFVLEHGFGEAFPOAFSGDPEGLMGMGCK
IGNYSCSYRLQAVPDINSFDSLELVGAAGVADECTGCVNIVRSTLLVYVVALGIV
LAIEWKMFKRLKTSGETDGRDPTVLEBCNDQDGRGATNTPSSALQGVSEVGTPI
"
```

BASE COUNT 1105 a 1123 c 1078 g 1101 t

ORIGIN

alignment_scores: quality: 93.50 length: 130

Ratio: 1.438 Gaps: 8
Percent Similarity: 50.000 Percent Identity: 34.615

alignment_block:

US-09-471-276-831 x AF322216 ..

Align seq 1/1 to: AF322216 from: 1 to: 4407

```

3 MetLeuValValPheLeuLeuLeuTrpGlyValThrTrp.....GI 16
|||||.....|
175 ATGCTTCGAGACTTCCTTTGCTCTTGGATTGGCTCAATCGGG 224
16 yProValThrGluAlaIatIlePheTyrgLutThrGln**SerLeuTrpA 33
|.....|
225 TATCACTTCATGAGCAGAGTG.....GAGTCGACACGAGAGCTATGCA 268
33 IagLysGlnHis**LeuLysThrLeuGlnGlnCysAspAlaAspVal 49
|||||.....|
269 TAGAGTCCAACTACCC..... 284
50 ProGlyProProGlyAspSerArg.....LeuProAl 60
|||||.....|
285 CCAGGCCCTTGGGAGAACATCACACTTGGTGCAAAAGCCCTCTCGGG 334
60 aValGlnGluTrpGlyAlaGlnGluProValHis.....LeuAspS 74
|.....|
335 TATCCACAGAGTTCTCTGCTGTGAAGATACACACAGATGACTGGATT 384
74 eP.ProAlaIleLys.....HisGlnPheLeu.....LeuTh 84
|||...|
385 CATTCTCTTACAGACCTTCACAGTTTCATTTTCATAGTGGCCCTTAC 434
84 rGlyAspThrGlnGlyArgTyrgCysArgSerGlyLeuSerThrGlyT 101
|.....|
435 TGAGTCCAAATACAGGCTTTTATGATGCTCTGCTACTGGAATGAGAAAGCT 484
101 rp***GlnLeuSerLysLeuLeuGluLeuThrGlyPro 113
|||||.....|
485 GGTGAGAGCCCAATAATTCGTACAGCTACAGCAGCA 522

```

seq_name: gb_fr:AF009643

seq_documentation_block:

LOCUS AF009643 1924 bp mRNA PRI 05-DEC-1997
DEFINITION Homo sapiens clone 6 immunoglobulin-like transcript 5 protein mRNA,
complete cds.

ACCESSION AF009643
VERSION AF009643.1 GI:2662445

KEYWORDS human.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 1924)

AUTHORS Colonna,M., Navarro,F., Bellon,T., Liano,M., Garcia,P.,
Samoridis,J., Angman,L., Cella,M. and Lopez-Botet,M.

TITLE A common inhibitory receptor for major histocompatibility complex
class I molecules on human lymphoid and myelomonocytic cells

JOURNAL J. Exp. Med. 186 (11), 1809-1818 (1997)

MEDLINE 98044246

REFERENCE 2 (bases 1 to 1924)

AUTHORS Colonna,M.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-1997) Basel Institute for Immunology, 487
Grenzacherstrasse, Basel CH-4005, Switzerland

FEATURES
source location/Qualifiers

```

1..1924
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="myelomonocytic"
/chromosome="19"
/clone="6"
/map:"19q13.4"

```



```

88 Gln.....GlyArgTyrArgCysArgSerGlyLeuSerThrGlyTr 101
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
270 CACGACCATCAGGAGATACCGCTGCCACTATTACAGCTCTGCAGGCTG 319
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
101 p***GlnLeuSerLysLeuLeuGluLeu 110
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
320 GTCAGACCCAGCAGCCGCCCTGGAGCTG 347

seq_name: gb_pr:AF009639

seq_documentation_block:
LOCUS      AF009639          2063 bp      mRNA      PRI      05-DEC-1997
DEFINITION Homo sapiens clone 31 immunoglobulin-like transcript 5 protein
ACCESSION  AF009639
VERSION    AF009639.1  GI:2662437
KEYWORDS   .
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 2063)
AUTHORS   Colonna,M., Navarro,F., Bellon,T., Llano,M., Garcia,P.,
            Samaridis,J., Angman,L., Cella,M. and Lopez-Botet,M.
TITLE     A common inhibitory receptor for major histocompatibility complex
            class I molecules on human lymphoid and myelomonocytic cells
JOURNAL    J. Exp. Med. 186 (11), 1809-1818 (1997)
MEDLINE    98044246
AUTHORS    2 (bases 1 to 2063)
            Colonna,M.
REFERENCE  Direct Submission
TITLE     Submitted (20-JUN-1997) Basel Institute for Immunology, 487
JOURNAL    Grenzachstrasse, Basel CH-4005, Switzerland
FEATURES   source
            1..2063
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /cell_type="myelomonocytic"
            /chromosome="19"
            /clone="31"
            /map="19q13.4"
            3..1898
            /note="immunoglobulin superfamily; inhibitory receptor"
            /codon_start=1
            /product="immunoglobulin-like transcript 5 protein"
            /protein_id="AAC51894.1"
            /db_xref="GI:2662438"
            /translation="MPALTAICGLSLGPRTRQAQPPFKPTLMAEPGVSISNGSP
            VTWCGSLAEAEYOLDKESPEPDNRNPLEPKKARFSLPSMTQHAGRYCHYIS
            SAGWSEPSDPLELVMTGYNKTLSALPSPVAVSGNMTLRGSGKGYHHFVLMKEIG
            HQLPRTLDSDQLHSGFQALPVPVPTPSHWRFTCYYYNTPMWMSHPSDPLEILP
            SGVSRKPSLTLQCPVLAAGOSLTLQCSVDGYDFVLYKEEDRFLDRPGQDPAGL
            SOANFTLPSVRSYGGYRCYGAHNLSEMSAPSDPLILITGOIYDVSLSAOPGPT
            VASGEMTLLOSRGYFPDPLITLKEGAHPRILSMKHYOAEFMSPTSAHAG
            TYRCGSRSSNPHILSPSEPLELAVSGHSGSSLPPIGPSTPGLGYLEVLIGVY
            AFVLLEFLFLPLLRKHSHKRTSDOKRTDQRPAGAETEPRKGLLRSSPADY
            QEENLYAAVKDQSEDRVLEDSQSDHEDPAVYAPVKHSSPREMSPPSSLSGEF
            LDTKRDQVEEDROMDTEAASEASQDVTYAOIHSILTRRKATEPPPSOEGEPAPPSI
            YATLAIH"
BASE COUNT  449 a      690 c      561 g      363 t
ORIGIN

```

alignment_scores:

Quality:	92.50	Length:	126
Ratio:	1.542	Gaps:	5
Percent Similarity:	47.619	Percent Identity:	30.159

alignment_block:

US-09-471-276-831 x AF009639 ..

Align seq 1/1 to: AF009639 from: 1 to: 2063

```

1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3 ATGACGCCGCCCTCAGAGCCCTGCTCTGCTGCTGGGTGAGTGTGGGCC 52
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
17 ovalThrGluAlaAlaIlePheTyrGluThrGln***SerLeuTrpAlaG 34
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
53 CAGACCCGCGATCAGCAGCAGGCCCTTCCCAAAACCAACCCCTCTGGGCTG 102
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
34 Lu.....SerLysHis***LeuLysThrLeu 42
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
103 AGCCAGGCTCTGTGATCAGCTGGGAGAGCCCTGACACCATCTGCTGTAG 152
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
43 GlyGlnCysAspAla.....AspValProGlyProGly 54
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
153 GGGAGCCTGAGGAGCCGAGAGTACCAACTGATTAAGAGGAGCAAGCCCA.. 200
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
54 yAspSerArgLeuProAlaValGlnGluTrpGlyAlaGlnGluProValH 71
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
201 .....GAGCCTGTGGAGAGAAATMAACCCACTG.. 227
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
71 tLeuAspSerProAlaIleLysHisGlnPheLeuLeuThrGlyAspThr 87
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
228 .....GAACCCAGACAGAGCCAGATTTCATCCATCCATCAGAGACA 269
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
88 Gln.....GlyArgTyrArgCysArgSerGlyLeuSerThrGlyTr 101
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
270 CACGACCATCAGGAGATACCGCTGCCACTATTACAGCTCTGCAGGCTG 319
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
101 p***GlnLeuSerLysLeuLeuGluLeu 110
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
320 GTCAGACCCAGCAGCCGCCCTGGAGCTG 347

seq_name: gb_pr:AF031555

seq_documentation_block:
LOCUS      AF031555          2063 bp      mRNA      PRI      09-DEC-1997
DEFINITION Homo sapiens clone 17.23 immunoglobulin-like transcript 5 mRNA,
ACCESSION  AF031555
VERSION    AF031555.1  GI:2665644
KEYWORDS   .
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 2063)
AUTHORS   Colonna,M., Navarro,F., Bellon,T., Llano,M., Garcia,P.,
            Samaridis,J., Angman,L., Cella,M. and Lopez-Botet,M.
TITLE     A common inhibitory receptor for major histocompatibility complex
            class I molecules on human lymphoid and myelomonocytic cells
JOURNAL    J. Exp. Med. 186 (11), 1809-1818 (1997)
MEDLINE    98044246
AUTHORS    2 (bases 1 to 2063)
            Colonna,M.
REFERENCE  Direct Submission
TITLE     Submitted (27-OCT-1997) Basel Institute for Immunology, 487
JOURNAL    Grenzachstrasse, Basel 4005, Switzerland
FEATURES   source
            1..2063
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="19q13.4"
            /chromosome="19"
            /cell_type="myelomonocyte"
            /clone="17.23"
            1..1898
            /note="inhibitory receptor; IL75"
            /codon_start=3
            /product="immunoglobulin-like transcript 5"
            /protein_id="AAC51902.1"
            /db_xref="GI:2665645"
            /translation="MPALTAICGLSLGPRTRQAQPPFKPTLMAEPGVSISNGSP
            VTWCGSLAEAEYOLDKESPEPDNRNPLEPKKARFSLPSMTQHAGRYCHYYS

```


HQLPRLDSQQLHSGFQALFPVGPVTPSPHMRFTCYYYTNTPMWMSHPDLELP
 SGVSRKPSLTLGGPVLAAGSLTLOGSDGVDFEFLYKGEERPELORPGOPAGL
 SQANFTLGPVSRSGGQRCYGAHNLSEMSRPSPLNIMAGQIYDVSLSAQGPT
 VASGEMNLTLCQSGYFDTFLYKEGAHNPRLRSMGAKRYQAEFPMSPTSAHAG
 TYRCYGRSSNPHLSPSEPLEMWSHSGSSSLPPGPPSPGLGRYLEVLIGVSV
 AFVLLFLFLFLRLRORSHKRTSDQKRTQFORPAGAAETPKRGILRRSPADY
 OENLYAAVKDQSEEDVLEDSQSDPHEDPAVLYAVVKHSSPREMASSPSLSGE
 FLDPKROVEEDRQMDTRAAASASQDVTYQALHSLTLRRATTEPPSPQEGEPAPES
 IYATLAIH"

BASE COUNT 450 a 692 c 563 g 361 t
 ORIGIN

alignment_scores:
 Quality: 92.50 Length: 126
 Ratio: 1.542 Gaps: 5
 Percent Similarity: 47.619 Percent Identity: 30.159

alignment_block:
 US-09-471-276-831 x AF009642 ..

Align seg 1/1 to: AF009642 from: 1 to: 2066

```

1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
  |||::: ||::: ||::: |||::: |||::: |||:::
3 ATGACGCGCCGCCCTGACAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 52
17 ovalThrGluAlaAlaIlePheTyrgIuThrGln**SerLeuTrpAlaG 34
  |||::: |||::: |||::: |||::: |||::: |||:::
53 CAGGACCCCGCATGCGAGGAGGCGCCCTTCCCAAAACCCACCTCGGGCTG 102
34 Lu.....SerGluHis**LeuLysThrLeu 42
  |||::: |||::: |||::: |||::: |||::: |||:::
103 AGCCAGGCTGTGTATCAGTGGGGAGGCCCGCTGACCATCTGGTGTGAG 152
43 GlyGlnCysAspAla.....AspValProGlyProProGly 54
  |||::: |||::: |||::: |||::: |||::: |||:::
153 GGGAGGCTGAGAGGCCCGAGAGTACCACTGATTAAGAGGGAACCCCA.. 200
54 yAspSerArgLeuProAlaValGlnGluTrpGlyAlaGlnGluProValH 71
201 .....GAGCCCTGGGACAGAAATATACCCACTG. 227
71 ILeuAspSerProAlaIleLysHisGlnPheLeuLeuThrGlyAspThr 87
  |||::: |||::: |||::: |||::: |||::: |||:::
228 .....GAACCCAGACAGAGCCGAGATTCTCCATCCATCCATCCATGACA 269
88 Gln.....GlyArgTyrgCysArgSerGlyLeuSerThrGlyTr 101
  |||::: |||::: |||::: |||::: |||::: |||:::
270 CAGCACCATGACGAGGAGATACCGCTGCCACTATTACAGCTCTGCGAGCTG 319
101 p***GlnLeuSerLysLeuLeuGlnLeu 110
  |||::: |||::: |||::: |||::: |||::: |||:::
320 GTCAGAGCCCGACGAGCCCGCTGAGCTG 347
```

seq_name: gb_pr.AF009636

seq_documentation_block:
 LOCUS AF009636 2091 bp mRNA PRI 05-DEC-1997
 DEFINITION Homo sapiens clone 17.8 immunoglobulin-like transcript 5 protein
 mRNA, complete cds.

ACCESSION AF009636
 VERSION AF009636.1 GI:2662431

KEYWORDS

SOURCE

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

TITLE

1 (bases 1 to 2091)
 Samanidis, J., Angman, L., Cella, M., and Lopez-Botet, M.
 A common inhibitory receptor for major histocompatibility complex
 class I molecules on human lymphoid and myelomonocytic cells

JOURNAL J. Exp. Med. 186 (11), 1809-1818 (1997)
 MEDLINE 98044246
 REFERENCE 2 (bases 1 to 2091)
 AUTHORS Colonna, M.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUN-1997) Basel Institute for Immunology, 487
 Grenzacherstrasse, Basel CH-4005, Switzerland

FEATURES
 source location/Qualifiers
 1..2091
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_type="myelomonocytic"
 /chromosome="19"
 /clone="17.8"
 /map="19q13.4"
 3..1472
 /note="Immunoglobulin superfamily; Inhibitory receptor"
 /codon_start=1
 /product="Immunoglobulin-like transcript 5 protein"
 /protein_id="AAC51891.1"
 /db_xref="GI:2662432"
 /translation="MPTALTLCLGLSLGPTTRMQAGPFPKPTLMAEBSVLSGSP
 VTIMCGSLAEQYLDKESPEMDRNRPLEPKRKARFISMTQHNHAGRYCHYVS
 SAGSEPSDPLELVTMGFYKNRPTLSALPSPVVASGNNMTLRGSGKGYHFLMKEGE
 HOLPRLDSQQLHSGFQALFPVGPVNSHMRFTCYYYTNTPMWMSHPDLELP
 SGVSRKPSLTLGGPVLAAGSLTLOGSDGVDFEFLYKGEERPELORPGOPAGL
 SQANFTLGPVSRSGGQRCYGAHNLSEMSRPSPLNIMAGQIYDVSLSAQGPT
 VASGEMNLTLCQSGYFDTFLYKEGAHNPRLRSMGAKRYQAEFPMSPTSAHAG
 TYRCYGRSSNPHLSPSEPLEMWSHSGSSSLPPGPPSPGLGRYLEVLIGVSV
 AFVLLFLFLFLRLRORSHKRTSDQKRTQFORPAGAAETPKRGILRRSPADY
 OENLYAAVKDQSEEDVLEDSQSDPHEDPAVLYAVVKHSSPREMASSPSLSGE
 FLDPKROVEEDRQMDTRAAASASQDVTYQALHSLTLRRATTEPPSPQEGEPAPES
 IYATLAIH"

CDS

BASE COUNT 452 a 702 c 570 g 367 t
 ORIGIN

alignment_scores:
 Quality: 92.50 Length: 126
 Ratio: 1.542 Gaps: 5
 Percent Similarity: 47.619 Percent Identity: 30.159

alignment_block:
 US-09-471-276-831 x AF009636 ..

Align seg 1/1 to: AF009636 from: 1 to: 2091

```

1 MetSerMetLeuValAlaPheLeuLeuLeuTrpGlyValThrTrpGlyPr 17
  |||::: ||::: ||::: |||::: |||::: |||:::
3 ATGACGCGCCGCCCTGACAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 52
17 ovalThrGluAlaAlaIlePheTyrgIuThrGln**SerLeuTrpAlaG 34
  |||::: |||::: |||::: |||::: |||::: |||:::
53 CAGGACCCCGCATGCGAGGAGGCGCCCTTCCCAAAACCCACCTCGGGCTG 102
34 Lu.....SerGluHis**LeuLysThrLeu 42
  |||::: |||::: |||::: |||::: |||::: |||:::
103 AGCCAGGCTGTGTATCAGTGGGGAGGCCCGCTGACCATCTGGTGTGAG 152
43 GlyGlnCysAspAla.....AspValProGlyProProGly 54
  |||::: |||::: |||::: |||::: |||::: |||:::
153 GGGAGGCTGAGAGGCCCGAGAGTACCACTGATTAAGAGGGAACCCCA.. 200
54 yAspSerArgLeuProAlaValGlnGluTrpGlyAlaGlnGluProValH 71
201 .....GAGCCCTGGGACAGAAATATACCCACTG. 227
71 ILeuAspSerProAlaIleLysHisGlnPheLeuLeuThrGlyAspThr 87
  |||::: |||::: |||::: |||::: |||::: |||:::
228 .....GAACCCAGACAGAGCCGAGATTCTCCATCCATCCATCCATGACA 269
88 Gln.....GlyArgTyrgCysArgSerGlyLeuSerThrGlyTr 101
  |||::: |||::: |||::: |||::: |||::: |||:::
270 CAGCACCATGACGAGGAGATACCGCTGCCACTATTACAGCTCTGCGAGCTG 319
101 p***GlnLeuSerLysLeuLeuGlnLeu 110
  |||::: |||::: |||::: |||::: |||::: |||:::
```

```

|||||: ||| |||||||
320 GTCAGACCCGACGCCCTTGAGCTG 347
seq_name: gb_pr:AF009640

seq_documentation_block:
LOCUS AF009640 2091 bp mRNA PRI 05-DEC-1997
DEFINITION Homo sapiens clone J3 Immunoglobulin-like transcript 5 protein
mRNA, complete cds.
ACCESSION AF009640
VERSION AF009640.1 GI:2662439
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Colonna,M., Navarro,F., Bellon,T., Llano,M., Garcia,P.,
Samartidis,J., Angman,L., Cella,M. and Lopez-Botet,M.
A common inhibitory receptor for major histocompatibility complex
class I molecules on human lymphoid and myelomonocytic cells
J. Exp. Med. 186 (11), 1809-1818 (1997)
JOURNAL 98044246
REFERENCE 2 (bases 1 to 2091)
AUTHORS Colonna,M.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-1997) Basel Institute for Immunology, 487
Grenzacherstrasse, Basel CH-4005, Switzerland
FEATURES
Source
1..2091
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/cell_type:"myelomonocytic"
/chromosome:"19"
/clone:"33"
/map:"19q13.4"
3..1472
/feature:"Immunoglobulin superfamily; Inhibitory receptor"
/codon_start:1
/product:"Immunoglobulin-like transcript 5 protein"
/protein_id:"ANC51895.1"
/db_xref:"GI:2662440"
/translation:"MTPALTYALLGLSLGSPRTMQAGPFRKPTLMARPGSVISNGSP
VTIWCOSLEAOEYOLDKESPEPDNRNPLEPKARFSTISMQHAGRYCHYYS
SAGSEPDPLELVMTGYNKPRTLSALSPYVAGCNMLKSGSKGHVFLMEKE
HOLPRTDSOOLHSGFOALPVPVGPVSHMRFRFYCYTPTPMVMSHSPLEILP
SGVSRKPSLITQGPVLAAGOSTLTQCCSDYVDYFVLYKEERFLRPGQPPACL
SOANFTLGPVSKSTGGYRCYGAHNLSSMSAPSDPLILITGQIYDVVSLSAOPCPT
VASGNMTLLCOSRGYFDYFTLTKEGAHPLRLRSMYGAHLYQAEFPMSPTSAHAG
TYRCYGRSSNPHLSPSEPLEMLVSGHSGSSLPPTGPPSTPGLGRYLEVLIGVSY
AFVILLFLLFLRRORSHKRTSGRLCPSSPRPEKO"
BASE COUNT 454 a 703 c 566 g 368 t
ORIGIN

alignment_scores:
Quality: 92.50 Length: 126
Ratio: 1.542 Gaps: 5
Percent Similarity: 47.619 Percent Identity: 30.159

alignment_block:
US-09-471-276-831 x AF009640 ..
Align seg 1/1 to: AF009640 from: 1 to: 2091
1 MetSerMetLeuValValPheLeuLeuLeuTrrpGlyValThrTrpGlyPr 17
|||||: ||||| |||||: |||||
3 ATGACGCCCGCCCTGACACGCCCTGCTGCTGCGCTGAGCTGCGGCC 52
|||||: ||||| |||||: |||||
17 ovalThrglAlaIalIelPheTrrgluThrgln**SerLeuTrpAla 34
|||||: ||||| |||||: |||||
53 CAGCAGCCGATCGACGAGGCCCTTCCCAAAACCCGCTCGCGCTG 102

```

```

34 Lu.....SerLuhis**LeuLysrhrLeu 42
||| ||| |||||
103 AGCCAGGCTCTGTGATCAGCTGGGAGGCCCTGACCATCTGTGTGAC 152
|||||: ||||| |||||: |||||
43 GlyGlnCysaspAla.....AspValProGlyProProG 54
|||||: ||||| |||||: |||||
153 GGGAGCCTGAGGAGCCGAGAGTACCACTGATTAACAGAGGAGCCCA.. 200
|||||: ||||| |||||: |||||
54 yAspSerArgLeuProAlaValGlnGluTrrpGlyAlaGlnGluProValH 71
|||||: ||||| |||||: |||||
201 .....GAGCCCTGGGACAGAAATACCCACTG. 227
|||||: ||||| |||||: |||||
71 ILeuAspSerProAlaIalIelYshISclnPhelLeuLeuThrglyAspThr 87
|||||: ||||| |||||: |||||
228 .....GAAACCAAGACAGAGCCGATTCATTCATCCATCCATCAGACA 269
|||||: ||||| |||||: |||||
88 Gln.....GlyArgTrrArgCysArgSerGlyLeuSerThrglyTr 101
|||||: ||||| |||||: |||||
270 CACACACCATCGACGAGATACCGCTGCCACTATTACACGTCTGACGCTG 319
|||||: ||||| |||||: |||||
101 P***GlnLeuSerLysLeuLeuGlnLeu 110
|||||: ||||| |||||: |||||
320 GTCAGACCCGACGCCCTTGAGCTG 347
seq_name: gb_pr:AC005590

seq_documentation_block:
LOCUS AC005590 4756 bp DNA PRI 02-SEP-1998
DEFINITION Homo sapiens chromosome 16, cosmid bridge clone 306E6 (LNL),
complete sequence.
ACCESSION AC005590
VERSION AC005590.1 GI:3513296
KEYWORDS HTG.
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Rieke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Melnick,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
Mitra,M. and Deaven,L.
Sequencing of Human Chromosome 16p13.3
JOURNAL Unpublished
AUTHORS Rieke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) system
JOURNAL Unpublished
AUTHORS Rieke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Melnick,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
and Deaven,L.
JOURNAL Submitted (02-SEP-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M886, Los
Alamos, NM 87545, USA
FEATURES
Source
1..4756
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/clone:"306E6"
/chromosome:"16"
/map:"16p13.3"
join(1..350,1825..2168)
/note:"97% identity EST z652b07.s1"
/db_xref:"dbEST:AN017253"
complement(join(2011..2173,3274..3514))
/note:"97% identity tmc43"
/db_xref:"dbEST:AA631976"
misc_feature
misc_feature
misc_feature

```



```

misc_feature complement(2022..2111)
              /note="GRAIL 2 excellent exon, frame 2"
misc_feature complement(join(2084..2173,3273..3512))
              /note="928 Identity mouse EST ms33a04.r1"
              /db_xref="dbEST:MA169004"
repeat_region 2956..3033
              /rpt_family="Alu"
              /note="GRAIL 2 excellent exon, frame 2"
BASE COUNT 915 a 1297 c 1466 g 1077 t 1 others
ORIGIN

alignment_scores:
  Quality: 92.50      Length: 107
  Ratio: 1.652      Gaps: 4
  Percent Similarity: 52.336      Percent Identity: 27.103

alignment_block:
  US-09-471-276-831 x AC005590/rev ..
  Align seg 1/1 to reverse of: AC005590 from: 1 to: 4756

10 LeuTPGly.....ValThrTPGlyProValThrGluAlaAla11 23
   :::::|||||
1420 GTGTGGGGGCGCAGAGTGTCTCCGCCGGCTTCAACCTGCCCTGCTCT 1371
   :::::|||||
23 ePeTYrGluThrGln***SerLeuTPAlaGluSerGluHis***LeuL 40
   :::::|||||
1370 TCTCTTCCACACAGCTCTCCCTCCCTCTGGGAAACTC. 1322
   :::::|||||
40 ysrhrLeuGlyGlnCysAspAlaAspValProGlyProProGlyAspSer 56
   ||| |||||
1321 .....TGCATGCTGCCCTGCCCTGCCCTGCCCTGCCATGCA 1286
   :::::|||||
57 ArgLeuProAlaValAlaGlnGluTPGlyAlaGlnGluProValHisLeuAs 73
   ||| :::::|||||
1285 CGTTCGCCAGCGCTGACCCAGTGTGGCTGC.....ATTCACTGGGA 1245
   :::::|||||
73 pSerProAlaIleLeuHisGlnPhe.....L 82
   :::::|||||
1244 GGGGCGCTGCCTCATCGGCGCTGCCATCCCTGCTCTTCTTGGCAGC 1195
   ||||| |||
82 euLeuThrGlyAspThrGlnGlyArgGlyArgGlySerGlyLeuSer 98
   ||||| |||
1194 TCTTCTCTGAAAGCTGAGGAGGACTTCTCTGCAAGGAGGACCCAA 1145
   :::::|||||
99 ThrGlyTrp***GlnLeuSer 105
   :::::|||||
1144 GTATTATGACACACTGACC 1124

seq_name: gb_hhg:AC041031

seq_documentation_block:
LOCUS AC041031 68207 bp DNA HTG 11-APR-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-684C2 map 2, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC041031.1 GI:7534234
VERSION AC041031.1 GI:7534234
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 68207)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Human sapiens chromosome 2, clone RP11-684C2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 68207)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

```

TITLE JOURNAL COMMENT

```

Collins,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardy,J., Grinde,S., Guyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Labrecque,K., Lamazares,R., Landers,T., Lehotzky,J.,
Levine,R., Liu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGuff,A., McKernan,K., McPheters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tittell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L9625
Center clone name: 684_C_2

* NOTE: This record contains 77 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
760 859: gap of 100 bp
860 1626: contig of 767 bp in length
1627 1726: gap of 100 bp
1727 2507: contig of 781 bp in length
2508 2607: gap of 100 bp
2608 3396: contig of 789 bp in length
3397 3496: gap of 100 bp
3497 4279: contig of 783 bp in length
4280 4379: gap of 100 bp
4380 5137: contig of 758 bp in length
5138 5237: gap of 100 bp
5238 5988: contig of 761 bp in length
5989 6098: gap of 100 bp
6099 6879: contig of 781 bp in length
6880 6979: gap of 100 bp
6980 7765: contig of 786 bp in length
7766 7865: gap of 100 bp
7866 8633: contig of 768 bp in length
8634 8733: gap of 100 bp
8734 9515: contig of 782 bp in length
9516 9615: gap of 100 bp
9616 10399: contig of 784 bp in length
10400 10499: gap of 100 bp
10500 11289: contig of 790 bp in length
11290 11389: gap of 100 bp
11390 12179: contig of 790 bp in length
12180 12279: gap of 100 bp
12280 13046: contig of 767 bp in length
13047 13146: gap of 100 bp

```

```
* 13147 13930: contig of 784 bp in length
* 13931 14030: gap of 100 bp
* 14031 14798: contig of 768 bp in length
* 14799 14898: gap of 100 bp
* 14899 15681: contig of 783 bp in length
* 15682 15781: gap of 100 bp
* 15782 16555: contig of 774 bp in length
* 16556 16655: gap of 100 bp
* 16656 17506: contig of 851 bp in length
* 17507 17606: gap of 100 bp
* 17607 18415: contig of 809 bp in length
* 18416 18515: gap of 100 bp
* 18516 19320: contig of 805 bp in length
* 19321 19420: gap of 100 bp
* 19421 20185: contig of 765 bp in length
* 20186 20285: gap of 100 bp
* 20286 21075: contig of 790 bp in length
* 21076 21175: gap of 100 bp
* 21176 21956: contig of 781 bp in length
* 21957 22056: gap of 100 bp
* 22057 22844: contig of 788 bp in length
* 22845 22944: gap of 100 bp
* 22945 23744: contig of 800 bp in length
* 23745 23844: gap of 100 bp
* 23845 24626: contig of 782 bp in length
* 24627 24726: gap of 100 bp
* 24727 25520: contig of 794 bp in length
* 25521 25620: gap of 100 bp
* 25621 26385: contig of 765 bp in length
* 26386 26485: gap of 100 bp
* 26486 27298: contig of 813 bp in length
* 27299 27398: gap of 100 bp
* 27399 28206: contig of 808 bp in length
* 28207 28306: gap of 100 bp
* 28307 29108: contig of 802 bp in length
* 29109 29208: gap of 100 bp
* 29209 30006: contig of 798 bp in length
* 30007 30106: gap of 100 bp
* 30107 30881: contig of 775 bp in length
* 30882 30981: gap of 100 bp
* 30982 31764: contig of 785 bp in length
* 31765 31864: gap of 100 bp
* 31865 32666: contig of 802 bp in length
* 32667 32766: gap of 100 bp
* 32767 33555: contig of 789 bp in length
* 33556 33655: gap of 100 bp
* 33656 34433: contig of 778 bp in length
* 34434 34533: gap of 100 bp
* 34534 35338: contig of 805 bp in length
* 35339 35438: gap of 100 bp
* 35439 36218: contig of 780 bp in length
* 36219 36318: gap of 100 bp
* 36319 37090: contig of 772 bp in length
* 37091 37190: gap of 100 bp
* 37191 37996: contig of 806 bp in length
* 37997 38096: gap of 100 bp
* 38097 38846: contig of 750 bp in length
* 38847 38946: gap of 100 bp
* 38947 39711: contig of 765 bp in length
* 39712 39811: gap of 100 bp
* 39812 40593: contig of 782 bp in length
* 40594 40693: gap of 100 bp
* 40694 41469: contig of 776 bp in length
* 41470 41569: gap of 100 bp
* 41570 42364: contig of 795 bp in length
* 42365 42464: gap of 100 bp
* 42465 43251: contig of 787 bp in length
* 43252 43351: gap of 100 bp
* 43352 44138: contig of 787 bp in length
* 44139 44238: gap of 100 bp
* 44239 45035: contig of 797 bp in length
* 45036 45135: gap of 100 bp
* 45136 45911: contig of 776 bp in length
```

```
* 45912 46011: gap of 100 bp
* 46012 46788: contig of 777 bp in length
* 46789 46888: gap of 100 bp
* 46889 47670: contig of 782 bp in length
* 47671 47770: gap of 100 bp
* 47771 48571: contig of 801 bp in length
* 48572 48671: gap of 100 bp
* 48672 49467: contig of 796 bp in length
* 49468 49567: gap of 100 bp
* 49568 50358: contig of 791 bp in length
* 50359 50458: gap of 100 bp
* 50459 51239: contig of 781 bp in length
* 51240 51339: gap of 100 bp
* 51340 52138: contig of 799 bp in length
* 52139 52238: gap of 100 bp
* 52239 53039: contig of 801 bp in length
* 53040 53139: gap of 100 bp
* 53140 53909: contig of 770 bp in length
* 53910 54009: gap of 100 bp
* 54010 54793: contig of 784 bp in length
* 54794 54893: gap of 100 bp
* 54894 55683: contig of 790 bp in length
* 55684 55783: gap of 100 bp
* 55784 56591: contig of 808 bp in length
* 56592 56691: gap of 100 bp
* 56692 57462: contig of 771 bp in length
* 57463 57562: gap of 100 bp
* 57563 58353: contig of 791 bp in length
* 58354 58453: gap of 100 bp
* 58454 59255: contig of 802 bp in length
* 59256 59355: gap of 100 bp
* 59356 60162: contig of 807 bp in length
* 60163 60262: gap of 100 bp
* 60263 61060: contig of 798 bp in length
* 61061 61160: gap of 100 bp
```

```
alignment_scores:
  Quality: 91.00      Length: 150
  Ratio: 1.379      Gaps: 8
  Percent Similarity: 44.000      Percent Identity: 24.667
```

alignment_block:

US-09-471-276-831 x AC041031/rev ..

Align seq 1/1 to reverse of: AC041031 from: 1 to: 68207

```
12 GlyValThrTrpGlyProValThrGlu..... 20
|||||.....
4995 GCGATGCTGGGATCCAGCCCTGATTCACATTAGAGCCCTCCCA 4946
21 .....AlaIAlaIlePheTyrGluThrGln* 29
4945 AACCCACCCCTGGGNTGAGCCAGGCTGTGTGATGAGCTGGGCGGCCG 4896
29 **SerLeuTrpAlaGluSerGlnHis**LeuLysThrLeuGlnCys 45
4895 TGACCATCTGTGTGTCAG.....GGAGGCTG 4870
46 AspAla.....AspValProGlyProGlyAspSerLeu 57
4869 GAGGCCGACGAGTACCAACTGCATTAAAGCGCAAGCCCA..... 4831
57 glauProAlaValGlnGluTrpGlyAlaGlnGluProValHisLeuAsp 74
4830 .....GAGCCCTGGGACAGAAATTAACCCCACTG.....G 4803
74 expProAlaIleLysHisGlnPheLeuLeuThrGlyAspThrGln..... 88
4802 AACCCAGAACAGACGACGATTCCTCCATCCATCCATGACAGACACCAT 4753
89 ...GlyArgTyrArgCysArgSerGlyLeuSerThrGlyTyr**GlnLe 104
4752 GCAGGAGATACCGCTCCACTATTACAGCTCTGCAGCGTGTACAGGCC 4703
```

```

seq_name: gb_pat:AX046972
seq_documentation_block:
LOCUS      AX046972          997 bp    DNA
DEFINITION Sequence 35 from Patent WO0068383.
ACCESSION  AX046972
VERSION     AX046972.1  GI:11876401
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 997)
AUTHORS     Cosman,D.,J., Anderson,D.M. and Borges,L.
TITLE        Family of immunoregulators designated leukocyte immunoglobulin-like
             receptors (ILr)
JOURNAL      Patent: WO 0068383-A 35 16-NOV-2000;
             IMMUNEX CORPORATION (US)
FEATURES
             Location/Qualifiers
             1..997
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             73..834
             /note="unnamed protein product"
             /codon_start=1
             /protein_id="CAC18939.1"
             /db_xref="GI:11876402"
             /translation="MAPWSHPASQLOPVGDVAPSPALMVLICGLNLSKATIMAPGSVY
             ISRNSYTIKCOGTLAEQEVLYKEGSPPEPMDTONPLEPKKARFSPISMTENHAGRR
             RCYRSPAGNPEPSDPLELVYGFNKPDTLSALPSPVYSGENVTLQCGSRLEPDRFL
             LTEGKGLKSWTLDLSOLPLPSGQFALPVGVPVPSHNMRLRCYSRHIIILQVSEPSD
             LLRLPVSGEENVYSSRTIQSGQTGCGELIKRQGSFC"
BASE COUNT 241 a      292 c      281 g      183 t
ORIGIN
alignment_scores:
Quality: 90.50          Length: 75
Ratio: 2.207           Gaps: 4
Percent Similarity: 54.667 Percent Identity: 37.333
alignment_block:
US-09-471-276-831 x AX046972 ..
Align seg 1/1 to: AX046972 from: 1 to: 997
45 CYSASPALAAspValProGlyProGlyAspSerArgLeuProAla.V 61
||||| ||| ||| ||||| ::|||::|
222 TGTGACCAATCGGCTGTCAAGGAGCCTGTGAGGCCACGAGATACCGTCTG 271
61 aGInGlu.....TTPGlyAlaGInGluProValHisLeu 72
||::||| ::|||::|||::|
272 TTTAAAGAGGAAAGCCCGAAGCCCTGGGACACACAGAACCCACGTG..... 315
73 AspSerProAlaIleLeuShsGInPheLeuLeuThnGlyAspPheGln.. 88
||||| ::||| ||| ::::|||::|
316 ...GAGCCCAAGACAAGAGCCACATTCCTCCATCCATCGACATGACGACA 362
89 .....GlyArgTyrArgCysArgSerGlyLeuSerThnGlyTyr**G 103
||||| ||||| ||||| ::|||::|||
363 CCATGCAAGGAGATACCGCTGTACTACTACACACCCCTGACAGCGCTGTAC 412
103 InLeuSerLysLeuGlnLeu 110
:: ||| ||||| |||||

```

```

seq_name: gb_pat:AX046968
413 AGCCGAGGAGCCCGTGAGCTG 435

seq_documentation_block:
LOCUS AX046968 1007 bp DNA PAT 15-DEC-2000
DEFINITION Sequence 31 from Patent WO0068383.
ACCESSION AX046968
VERSION AX046968.1 GI:11876397
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1007)
Cosman,D.J., Anderson,D.M. and Borges,L.
Family of immunoregulators designated leukocyte immunoglobulin-like
receptors (11r)
Patent: WO 0068383-A 31 16-NOV-2000;
JOURNAL IMMUNEX CORPORATION (US)
FEATURES
source
Location/Qualifiers
1..1007
/oranism="Homo sapiens"
/db_xref="taxon:9606"
95..958
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC18937.1"
/db_xref="GI:11876398"
/translation="MAPMSHPASLOLPVGGDPAVSPALMVLICIGNLISKATLMAEPGSV
ISRSNVTIRCOGLLEAOEYLYKESPEPDNDONLEPKKKAFSLPSMTEHHAGVY
RCYIYSPAGNSEPSDLELYVTGRTKFTLSALDSPVYTSENTLDCGSLRDRIT
LTREGDKLSTWLTQSILTPSGQFQALPRVGGVTSRHHMLKCYGSRHHLLQWNSPESD
LLELPSCAADNLSPSQNSDSQTSFALSHLDYAVENLIRMGAGLLVLVLLIFQDMH
SQRSPQAAAGR"
BASE COUNT 230 a 303 c 276 g 198 t
ORIGIN

alignment_scores:
Quality: 90.50 Length: 75
Ratio: 2.207 Gaps: 4
Percent Similarity: 54.667 Percent Identity: 37.333

alignment_block:
US-09-471-276-831 x AX046968 ..

Align seg 1/1 to: AX046968 from: 1 to: 1007

45 CysAspAlaAspValProGlyProProGlyAspSerArgLeuProAla.V 61
||||| ||| ||| ||||| ::|||::|
244 TGYGACCATCCGGGTGTCAGGGGACCTCGGAGGCCGAGATATACCGTGG 293
61 a|c|n|g|u.....T|p|c|y|a|c|n|g|u|p|p|o|v|a|l|h|s|l|e|u 72
||||| ||| ::|||::|
294 TTAAGGAGGAGAACCCAGAACCTCGGAGACACAGAACCCACATG..... 337
73 AspSerProAlaIleLeuYHisGlnPheLeuLeuThGlyAspThrGln.. 88
:::||| ||| ::|||::| ::| ::| ::| ::| ::| ::|
338 ...GAGCCCAAGAACAGGCGCATATTCATCCATCCATCCATGACAGAGA 384
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
89 .....G|A|T|G|T|A|T|G|C|y|S|a|r|g|S|e|r|G|y|L|e|u|S|e|r|Th|G|Y|T|T|P|**G 103
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
385 CCATGCGAGGAGATACCGGTACTACTACACACCTCGACGCTGCTGACG 434
103 In|u|S|e|r|y|S|l|e|u|G|u|L|e|u 110
::| ||| ||||| |||
435 AGCCGAGGAGCCCGTGAGCTG 457

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2002, 16:46:53 ; Search time 31.95 seconds

(without alignments)
576.849 Million cell updates/sec

Title: US-09-471-276-831

Perfect score: 661
Sequence: 1 MSMLVFLLMGVTMGVPE.....LLETGPKVLACSLALDGAS 126

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database :

1: SP:REMBL_17:*
2: SP:archaea:*
3: SP:bacteria:*
4: SP:fungi:*
5: SP:human:*
6: SP_invertebrate:*
7: SP_mhc:*
8: SP:organelle:*
9: SP:phage:*
10: SP:plant:*
11: SP:rodent:*
12: SP:virus:*
13: SP:vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130.5	19.7	513	11 09EPH1	09eph1 ratu
2	90.5	13.7	631	4 075022	075022 homo sapien
3	87.5	13.2	631	4 015471	015471 homo sapien
4	84.5	12.8	448	4 013468	013468 homo sapien
5	81.5	12.3	590	4 075021	075021 homo sapien
6	78.5	11.9	652	4 075025	075025 homo sapien
7	78.5	11.8	192	12 079089	079089 human immun
8	76	11.5	1327	4 015070	015070 homo sapien
9	75.5	11.4	244	2 09ZHP2	09zhp2 nostoc sp.
10	75.5	11.4	635	11 055002	055002 mus musculu
11	75.5	11.3	192	12 09Q719	09q719 homo sapien
12	75	11.3	458	4 09NVR4	09nvr4 homo sapien
13	75	11.3	458	4 09BVM3	09bvm3 homo sapien
14	75	11.3	647	4 09P209	09p209 homo sapien
15	75	11.3	648	4 09P209	09p209 mus musculu
16	74.5	11.3	680	11 055001	055001 mus musculu
17	74.5	11.3	192	12 09WPD4	09wpd4 human immun
18	74.5	11.3	192	12 073426	073426 human immun
19	74	11.2	192	12 073426	073426 human immun

20	74	11.2	289	4 075018	075018 homo sapien
21	74	11.2	650	4 075024	075024 homo sapien
22	72.5	11.0	187	10 09LVK0	09lvk0 arabidopsis
23	72.5	11.0	192	12 079076	079076 human immun
24	72.5	11.0	244	2 09ZHP0	09zhp0 nostoc sp.
25	72	10.9	104	12 073438	073438 human immun
26	72	10.9	192	12 073410	073410 human immun
27	72	10.9	192	12 073413	073413 human immun
28	72	10.9	192	12 073439	073439 human immun
29	72	10.9	192	12 091QD9	091qd9 human immun
30	72	10.9	475	5 09VNR2	09vnr2 drosophila
31	71.5	10.8	489	2 09HUR4	09hur4 pseudomonas
32	71	10.7	192	12 073412	073412 human immun
33	71	10.7	192	12 073435	073435 human immun
34	71	10.7	192	12 099B78	099b78 human immun
35	71	10.7	288	7 046780	046780 bos taurus
36	71	10.7	643	5 045087	045087 caenorhabdit
37	70	10.6	147	12 073436	073436 human immun
38	70	10.6	192	12 073434	073434 human immun
39	70	10.6	192	12 09DP29	09dp29 human immun
40	70	10.6	192	12 099450	099450 human immun
41	69.5	10.5	728	12 09YQY0	09yqy0 ranid herpe
42	69.5	10.5	841	11 054999	054999 mus musculu
43	69	10.4	192	12 071262	071262 human immun
44	69	10.4	192	12 071966	071966 human immun
45	69	10.4	192	12 071971	071971 human immun
46	69	10.4	192	12 073411	073411 human immun
47	69	10.4	192	12 073440	073440 human immun
48	69	10.4	192	12 079098	079098 human immun
49	69	10.4	192	12 079115	079115 human immun
50	69	10.4	192	12 079128	079128 human immun
51	69	10.4	192	12 080237	080237 human immun
52	69	10.4	192	12 011977	011977 human immun
53	69	10.4	254	12 088704	088704 porcine rep
54	69	10.4	324	5 09YOD2	09yod2 hypera post
55	69	10.4	458	11 09DP33	09dp33 mus musculu
56	69	10.4	489	4 075019	075019 homo sapien
57	68.5	10.4	244	10 09FOD2	09fod2 arabidopsis
58	68.5	10.4	831	11 09ER72	09er72 mus musculu
59	68.5	10.4	831	11 09ER68	09er68 mus musculu
60	68	10.3	185	12 073433	073433 human immun
61	68	10.3	192	12 073431	073431 human immun
62	68	10.3	192	12 073432	073432 human immun
63	68	10.3	192	12 078242	078242 human immun
64	68	10.3	192	12 079079	079079 human immun
65	68	10.3	192	12 079096	079096 human immun
66	68	10.3	192	12 089450	089450 human immun
67	68	10.3	192	12 012159	012159 human immun
68	68	10.3	192	12 055945	055945 human immun
69	68	10.3	192	12 055969	055969 human immun
70	68	10.3	192	12 09WPD8	09wpd8 human immun
71	68	10.3	192	12 09WPD9	09wpd9 human immun
72	68	10.3	192	12 09WPD9	09wpd9 human immun
73	68	10.3	192	12 09WPD9	09wpd9 human immun
74	68	10.3	192	12 09WPD9	09wpd9 human immun
75	68	10.3	192	12 09WPD9	09wpd9 human immun
76	68	10.3	192	12 09WPD9	09wpd9 human immun
77	68	10.3	192	12 09WPD9	09wpd9 human immun
78	68	10.3	192	12 09WPD9	09wpd9 human immun
79	68	10.3	192	12 09WPD9	09wpd9 human immun
80	68	10.3	192	12 09WPD9	09wpd9 human immun
81	68	10.3	192	12 09WPD9	09wpd9 human immun
82	68	10.3	192	12 09WPD9	09wpd9 human immun
83	68	10.3	192	12 09WPD9	09wpd9 human immun
84	68	10.3	192	12 09WPD9	09wpd9 human immun
85	67.5	10.2	118	4 09POT0	09pot0 homo sapien
86	67.5	10.2	141	4 09UPX1	09upx1 homo sapien
87	67.5	10.2	436	2 09X8H4	09x8h4 streptomyc
88	67.5	10.2	672	2 09K6M7	09k6m7 bacillus ha
89	67.5	10.2	2967	12 041892	041892 hepatitis g
90	67	10.1	192	12 073420	073420 human immun
91	67	10.1	192	12 073425	073425 human immun
92	67	10.1	192	12 073426	073426 human immun

OY	48	DVPGPCGSRPLPAQWEMOAEPHILDSPAIKHGCLLNGDPGRGYRCRSGSLGTGXOLSKL	107
Db	61	DKEGLPMARKQNQLPEQAKMAKFIPISTVY-----DSAGRCRYETRPACHSEPSDP	112
OY	108	LEL--TG----PKYLA	117
Db	113	LELVATGFYAEPFLTA	128

RESULT	7	
075025		
ID	075025	PRELIMINARY; PRT; 652 AA.

DT 01-NOV-1998 (TReMBuRel. 08, Created)
DT 01-NOV-1998 (TReMBuRel. 08, Last sequence update)
DT 01-JUN-2001 (TReMBuRel. 17, Last annotation update)
DE LEBECOCYTE IMMUNOGLOBULIN-LIKE RECEPTOR-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eulhelia; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cosman D., Fanger N., Borges L., Kubin M., Chin W., Peterson L.,
RA Hew M.L.;
RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
CC EMBL: AF009221; AAB63522.1; -.
DR HSSP: P43626; INKR.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_3
DR SMART: SM00409; Ig_2.
DR SMART: SM00410; Ig_Like; 1.
SQ SEQUENCE 652 AA; 7108 MW; 93AF8F021A2949D CRC64;

Query Match	11.98; Score 78.5; DB 4; Length 652;
-------------	--------------------------------------

Best Local Similarity 24.6%; Pred. No. 2.1;
Matches 41; Conservative 16; Mismatches 57; Indels 53; Gaps 7.

```

Oy      1 MSMLVFFLLLCVGTWCPVTEMAIYETQXSLMAESEHXLKTLGQCDADVPGPCCGSRLLPA 600
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      1 MTPILTVLICGLSLGPRTHVQAGHLKPTLMAEPGSI-----TQGGPVTLLRCQG 51

```

```

QY 61 VQE-----WGADPPVHDSPAIKHOF--LTGDTGRYRCRSGSLST-GWX 102
   ||      | | | | | | | | | | | | | | | | | | | | | |
Db 52 GQETQRYLYREKKTALWITRIPEQL---VKKGGFPPIPSITWEHAGRYRCYGGSDTAGRS 108

```

```

QY      103 QLSKLELT-----GPKVLAC--SLALDGS 126
      : | ||| | : | | : | | |
Db      109 ESSDPLELVVTGAYIKPTLSAQSPVNVSGCNVILQCDSDQVAFDGS 155

```

RESULT	8	
ID	Q79089	PRELIMINARY;
	Q79089	PRT; 192 AA

DT	01-NOV-1996 (TREMBLER). 01, Created)
DT	01-NOV-1996 (TREMBLER). 01, last sequence update)
DT	01-JUN-2001 (TREMBLER). 17, last annotation update,
DE	VIF (VIRAL INFECTIVITY FACTOR).
DE	VIF (VIRAL INFECTIVITY FACTOR).

OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus
OX NCBI_TaxID=11676;
an

RP SEQUENCE FROM N.A.
RC STRAIN-PATENT A30;
RX MEDLINE=94303241; Pubmed=8030283;
RA Wieland U., Hartmann J., Suhr H., Salzberger B., Biggers H.J.,

RA Kuehn J.E.: "In vivo genetic variability of the HIV-1 *wif* gene.";
RT Virology 203:43-51(1994).
RL EMBL, Z30601: CAB83078.1.
DR EMBL, JPR000475: Viral_infect.
DR InterPro: IPR000475: Viral_infect.
DR Pfam: PF00559: Vif; 1.
DR PRINTS: PR00349: VIRIONINFECT.
DR ProDom: P000063: Viral_infect. 1.
SO SEQUENCE 192 AA: 22676 MW; 3BB650CCB37C104D CRC64

Query Match 11.8%; Score 78; DB 12; Length 192;
Best Local Similarity 23.0%; Pred. No. 0.64;
Matches 23; Conservative 16; Mismatches 37; Indels 24; Gaps 3

```

Oy      8 LLLMGV-----TWCPVTEAALFETQXSLMAESEHXKLTLCQCDADVPGPQDSRLPAV 61
      :: | | : : : : | | : : | | : |
Db      8 MIVVQYDRLRIKFTWKSLSVKHHIYSRKATGNYRNHHESTHPRVSSVEHILPLGDARLYIT 67

```

```

QY      62 QENGAQAEFVHLDSPAIKHFLLTGDTQGRYRCRSLSTCW 10
      ||      ||| : : : ||| |
DB      68 TYWG-----LTGERE--WHLGQVSIEW 89

```

RESULT	9	
015070		
ID	015070	PRELIMINARY;
		PRT; 1327 AA

DT	01-JAN-1998	(TREMBlrel. 05, Created)
DT	01-JAN-1998	(TREMBlrel. 05, Last sequence update)
DT	01-JUN-2001	(TREMBlrel. 17, Last annotation update)
DE	K1100364	

GN KIAA0364.
O5 Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo

OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=BRAIN;

RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.:
RT "Prediction of the coding sequences of unidentified human genes. VIII.
RA MEDLINE:9/349984; PubMed:3205841;

CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
RI line complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).

```

CC      DOMAIN:
DR      EMBL; AB002362; BAA20819.1; -
DR      HSSP; P43626; 1NKR.
DR      InterPro; IPR003598; 1q_c2.

```

```
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_12.
DR SMART; SM00408; IGC2; 1.
```

Query Match 11.58; Score 76; DB 4; Length 1327;

23 IFYETQXSLAESEHXLKTLGGCCADVPCCPPGDSRLPAVQEMCAQEPVHLDS----- 74

```

Db 23 LMDDPELWIESNY-----PQAPWEN----ITLW-CRSPSRISSKFLLLKDK 655
07 75 -----PAIKH---QFL--LTGDTQGRYKCRSGJSTGWOJLSKLELTGP 113

```

Db 66 TQMTWIRPSHKTFQVSPFLIGALTESNAGLYRCYWKETGWSKPSKVLLEAP 117


```

RESULT 10
O9ZHP2 PRELIMINARY: PRT: 244 AA.
ID 09ZHP2:
AC 09ZHP2:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NSPI RESTRICTION ENDONUCLEASE.
GN NSPIR.
OS Nostoc sp. ATCC 29411.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=82765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC29411;
RX MEDLINE-99077292; PubMed-9862476;
RA Xu S.-Y., Xiao J.-P., Etzwiler L., Holden M., Alliotta J., Poh C.L.,
RA Dalton M., Robinson D.P., Petronzio T.R., Moran L., Ganatra M.,
RA Ware J., Slanko B., Benner J.;
RT "Cloning and expression of the Apali, Nspl, NsplI, SacI, Scal, and
RT Sapi restriction-modification systems in Escherichia coli.";
RL Mol. Genet. 260:226-231(1998).
DR EMBL: AF056036; AAC97191.1; -.
DR REBASE: 1391; Nspl.
KW Endonuclease.
SO SEQUENCE 244 AA; 28073 MW; F59EF95917A57AA6 CRC64;

```

```

Query Match 11.4%; Score 75.5; DB 2; Length 244;
Best Local Similarity 39.3%; Pred. No. 1.6;
Matches 22; Conservative 5; Mismatches 18; Indels 11; Gaps 4;
OY 55 DSLRPAVQEMGAQEPVHLS---PAIKHOF-ILTGDTGGRYR-----CRSGLTG 100
DB 132 DHRFP-MERWGASEPPLTSMNDNEIKRKFQLLKDTSGHNLKRSRCRCIKTG 186

```

```

RESULT 11
O55002 PRELIMINARY: PRT: 635 AA.
ID 055002:
AC 055002:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PAIRED-IG-LIKE RECEPTOR A11 (KILLER CELL INHIBITORY RECEPTOR-LIKE
DE PROTEIN P91C) (FRAGMENT).
GN PIRAI1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-129/SV;
RA Yamashita Y., Fukuta D., Tsuji A., Nagabukuro A., Matsuda Y.,
RA Nishikawa Y., Ohya Y., Ohmori H., Ono M., Takai T.;
RL J. Biochem. 123:0-0(1998).
CC -!- SIMILARITY: NO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
DR EMBL: AF041036; AAB96928.1; -.
DR HSSP: P43626; INKR.
DR MGI: 1276541; PIRAI1.
DR MGD: MGI:1276541; PIRAI1.
DR InterPro: IPR003599; I9.
DR InterPro: IPR003600; I9_Like.
DR InterPro: IPR003006; I9_MHC.
DR Pfam: PF00047; I9; 5.
DR SMART: SM00409; IG; 3.
DR SMART: SM00410; IG_Like; 2.
FT NON TER 635
FT SEQUENCE 635 AA; 70558 MW; 8A6CA830D1BDB3D9 CRC64;

```

Query Match 11.4%; Score 75.5; DB 11; Length 635;

```

Best Local Similarity 28.1%; Pred. No. 4.5;
Matches 32; Conservative 12; Mismatches 45; Indels 25; Gaps 5;
OY 6 VFLLMGVWGVTEA---AIFYEQSLMAESEHXLTGQCADV--PPPGDSRP 59
DB 215 VELLVGNLQKPLTIAEPGSAVIAKRAMTWCQ-----GNLAEVYFLNNGSKTK 266
OY 60 AVQEMGAQEPVHLSPAIKHOFILTGDTQ---GRYCRSGLTGMYXLSKLEL 110
DB 267 STGT-----LQDPGNKGRFFIPSVTOAHGQIRCYCSSAGNSQSPDTLEL 312

```

```

RESULT 12
O90719 PRELIMINARY: PRT: 192 AA.
ID 090719:
AC 090719:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE VIF.
GN VIF.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V1991;
RA Janssens W., Laukkanen T., Salminen M.O., Carr J.K.,
RA Van der Auwera G., Heyndrickx L., Van der Groen G., McCutchan F.E.;
RT "Hiv-1 subtype H near-full genome, reference strains and analysis of
RT subtype-H-containing inter-subtype recombinants.";
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF190127; AAF18396.1; -.
DR InterPro: IPR000475; Viral_infect.
DR Pfam: PF00559; VIF; 1.
DR PRINTS: PR00349; VIRIONINFECT.
DR ProDom: PD000063; VIRIONINFECT.
SO SEQUENCE 192 AA; 22589 MW; E9015AE1C24D5375 CRC64;

```

```

Query Match 11.3%; Score 75; DB 12; Length 192;
Best Local Similarity 21.6%; Pred. No. 1.4;
Matches 22; Conservative 15; Mismatches 41; Indels 24; Gaps 3;
OY 8 LILMGV-----TWGPVTEAIFYEQSLMAESEHXLTGQCADVPPGDSRLPAY 61
DB 8 MIVQYDVRMRIRKTNLSVKKHMTVSKKAKKMYRHHYESTNPRTSSVHPIVGDARLVIT 67
OY 62 QEWGAQEPVHLSPAIKHOFILTGDTQGRYCRSGLTGMYXQ 103
DB 68 TYWG-----LHTGERD--WHLGHGVSIEWRQ 91

```

```

RESULT 13
O9NVR4 PRELIMINARY: PRT: 458 AA.
ID 09NVR4:
AC 09NVR4:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE CDNA FLJ10565 FIS, CLONE NT2RP2002954.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";

```


RC STRAIN-B10.A; TISSUE=PERITONEUM;
 RX MEDLINE-98218758; PubMed-9538215;
 RA Yamashita Y., Fukuta D., Tsuji A., Nagabukuro A., Matsuda Y.,
 RA Nishikawa Y., Ohkuma Y., Ohmori H., Ono M., Takai T.,
 RT "Genomic structures and chromosomal location of p91, a novel murine
 RT regulatory receptor family."
 RL J. Biochem. 123:358-368(1998).
 CC -I- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN
 CC EMBL: AF055896; AAC40073.1; -.
 DR HSSP: P43626; INKR.
 DR MGD: MGI:1276541; P1ra11.
 DR InterPro: IPR003599; Iq.
 DR InterPro: IPR003600; Iq_like.
 DR InterPro: IPR003006; Iq_MHC.
 DR Pfam: PF00047; Iq; 5.
 DR SMART: SM00409; Iq; 3.
 DR SMART: SM00410; Iq_like; 2.
 FT NON_TER
 FT SEQUENCE 663 AA; 73761 MW; 6E98B6E10BA8BB95 CRC64;

Query Match 11.3%; Score 74.5; DB 11; Length 663;
 Best Local Similarity 28.6%; Pred. No. 6.1;
 Matches 32; Conservative 11; Mismatches 48; Indels 21; Gaps 4;
 QY 6 VELLMGVTVGPVTEA----AIFYETQXSLMAESEXHLTLGGCDADVPDPPDSRLP 61
 DB 215 VELLVSGNLQKPTIKAEPSVITSKRAMTIWCQ-----GNLDAEV-----YFLHNE 260
 QY 62 QEWGAQEPVHLDSPAIKHQFLTLGDTQ---GRYRCRSGSLSTGXOLSKLLEL 110
 DB 261 KSKQTKSTQTLQDPGNKGRFFIPSVTQNAAGYRCYCYSAGMSQPSDTLEL 312

RESULT 18
 055001 PRELIMINARY: PRT: 680 AA.
 AC 055001:
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PAIRED-IG-LIKE RECEPTOR A10 (KILLER CELL INHIBITORY RECEPTOR-LIKE
 DE PROTEIN P91B).
 GN PIRA10.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId-10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV;
 *RA Yamashita Y., Fukuta D., Tsuji A., Nagabukuro A., Matsuda Y.,
 RA Nishikawa Y., Ohkuma Y., Ohmori H., Ono M., Takai T.,
 RL J. Biochem. 123:0-0(1998).
 CC -I- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 CC EMBL: AF041035; AAB96927.1; -.
 DR HSSP: P43626; INKR.
 DR MGD: MGI:1276542; P1ra10.
 DR InterPro: IPR003599; Iq.
 DR InterPro: IPR003600; Iq_like.
 DR InterPro: IPR003006; Iq_MHC.
 DR Pfam: PF00047; Iq; 5.
 DR SMART: SM00409; Iq; 3.
 DR SMART: SM00410; Iq_like; 2.
 SQ SEQUENCE 680 AA; 75597 MW; 5311DFE6D291BF11 CRC64;

Query Match 11.3%; Score 74.5; DB 11; Length 680;
 Best Local Similarity 28.6%; Pred. No. 6.2;
 Matches 32; Conservative 11; Mismatches 48; Indels 21; Gaps 4;

QY 6 VELLMGVTVGPVTEA----AIFYETQXSLMAESEXHLTLGGCDADVPDPPDSRLP 61
 DB 215 VELLVSGNLQKPTIKAEPSVITSKRAMTIWCQ-----GNLDAEV-----YFLHNE 260
 QY 62 QEWGAQEPVHLDSPAIKHQFLTLGDTQ---GRYRCRSGSLSTGXOLSKLLEL 110
 DB 261 KSKQTKSTQTLQDPGNKGRFFIPSVTQNAAGYRCYCYSAGMSQPSDTLEL 312
 RESULT 19
 09WPD4 PRELIMINARY: PRT: 192 AA.
 ID 09WPD4:
 AC 09WPD4:
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE VIF PROTEIN.
 GN VIF.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxId-11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-V108003A1L;
 RA Hassane G., Agostini I., Candotti D., Bessou G., Caballero M.,
 RA Agut H., Autran B., Barthalay Y., The French ALT Study Group,
 RA Vigne R.,
 RT "Characterization of human immunodeficiency virus type 1 vif gene in
 RT long-term asymptomatic individuals."
 RT Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 RL EMBL: AF143108; AAD37875.1; -.
 DR EMBL: AF143108; AAD37875.1; -.
 DR InterPro: IPR000475; Viral_infect.
 DR Pfam: PF00559; Vif; 1.
 DR PRINTS: PR00349; VIRIONINFECT.
 DR ProDom: PD000063; Viral_infect; 1.
 DR SEQUENCE 192 AA; 22805 MW; CEB7D94BB76F6280 CRC64;

Query Match 11.2%; Score 74; DB 12; Length 192;
 Best Local Similarity 22.0%; Pred. No. 1.8;
 Matches 22; Conservative 15; Mismatches 39; Indels 24; Gaps 3;
 QY 8 LLLMGV-----TWGVTEAIFYETQXSLMAESEXHLTLGGCDADVPDPPDSRLP 61
 DB 8 MIVQYDRMRIRITKXSLVKNHIIYSKAKOVYRHHYESTHPRISSVHPLGDARLVIT 67
 QY 62 QEWGAQEPVHLDSPAIKHQFLTLGDTQGRYRCRSGSLSTGW 101
 DB 68 TYWG-----LHTGERE--WHLGQGVSIEM 89

RESULT 20
 075018 PRELIMINARY: PRT: 289 AA.
 ID 075018:
 AC 075018:
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE LEUCOCYTE IMMUNOGLOBULIN-LIKE RECEPTOR-6B.
 GN LIR-6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId-9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Borges L., Hsu M.-L., Fanger N., Rubin M., Cosman D.,
 RA J. Immunol. 0:0-0(1997).
 CC -I- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 CC EMBL: AF025529; AAB87663.1; -.
 DR HSSP: P43626; INKR.
 DR InterPro: IPR003600; Iq_like.

DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 2.
DR SMART: SMO0410; Ig_Like; 1.
SQ SEQUENCE 289 AA; 3123 MW; E1930C4CEBAEA119 CRC64

Query Match	11.28;	Score 74;	DB 4;	Length 289;
Best Local Similarity	27.78;	Pred. No. 2.8;		
Matches 36;	Conservative 15;	Mismatches 45;	Indels 34;	Gaps 7

OY	1	MSMLVFLLINGVINGVPTVEAAIIEYEOXSLMASEHXKLTKG	-----CDA-----	47
		! : : : : ! : : : ! : ! ! ! !		
Db	1	MTPLTFVLCIGLSIGPRTHVQACLTLPKPTLMAPGCVI	-TQSPVITMCGILLETQYR	59
OY	48	---DVPQPGSRILPANGVEMGAEVILLDSATIKHOF	---LTGDDTGRCRCRGSLT	G 100
		! : : : : ! : : : ! : : : ! : ! ! ! !		
Db	60	LYREKKTAPWITRIP	-----QELVK-----KGQFIPISITWHTGRTRCFGSHTAG	106
OY	101	WXSLSKLEEL	110	
		! : : : ! ! !		
Db	107	WSPDPLPLEL	116	

RESULT	21	
075024		
ID	075024	PRELIMINARY;
		PRT; 650 AA

DT 01-NOV-1998 (TREMBLrel_08, Created)
 DT 01-NOV-1998 (TREMBLrel_08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel_17, Last annotation update)
 DE LEUCOCYTE IMMUNOGLOBULIN-LIKE RECEPTOR-1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID:9606;

17) SEQUENCE FROM N.A.
 RP Cosman D., Fanger N., Borges L., Kubin M., Chin W., Peterson L.,
 RA Hsu M.L.;
 RL Submitted (JUL-1997) to the EMBL/Genbank/DDBJ databases.
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN
 DR EMBL: AF0092220; AAB63521.1; -;
 DR HSSP: P43626; INKR.
 DR InterPro: IPR003599; Ig.
 DR InterPro: IPR003600; Ig-like.
 DR InterPro: IPR003006; Ig_MHC.
 DR Pfam: PF000647; Ig_4;
 DR SMART: SM00409; IG_2;
 DR SMART: SM00410; IG-like; 1.
 DQ SEQUENCE 650 AA; 70816 MW; D1321C0982B44BCA CRC64;

Query Match	11.28;	Score 74;	DB 4;	Length 650;
Best Local Similarity	26.68;	Pred. NO. 6.8;		
Matches	34;	Conservative	14;	Mismatches 50;
			Indels	30;
			Gaps	5

```

QY      1 MSMLVEFLLLGVMGAPVTEAAIIEYEQXSLAMASEHXILKTLGGCADVGPEDSDSLPA 60
      2 ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      1 MTPILTVLIGLGLGSPRTVQAGHLKPKPLTMAEBSYLT-----TGGSPVTLKCGG 51
QY      61 VQE-----MGADPEPHLDSPARKNQF---LTLGDTQGRYRCRSGLST-GMX 102
      2 ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      52 GGEQDEQENLYREKKTAPMTIRIPDEL--YKGGQPIPSITWEHAGRTCYGSDTAGRS 108
QY      103 QLSKLLLEL 110
      2 ::::: :::::
Db      109 ESSDPLEL 116
      2 :::::

```

RESULT	22	
Q9LVK0		
ID	Q9LVK0	PRELIMINARY; PRT; 187 AA

AC 09LVK0:2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE GB|AAC36161.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryotes; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eumastids; II: Brassicales; Brassicaceae; Arabidopsis.
 OX NCBL_TaxID:3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE:20277480; PubMed:10819329;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
 RT clones.";
 RL DNA Res. 7:131-135(2000).
 EMBL: AB019229; BAB02330.1; -.
 Q0 SEQUENCE 187 AA; 20916 MW; 41B9AE4E9086299 CRC64;

Query Match	11.0%	Score 72.5	DB 10	Length 187
Best Local	Similarly	25.9%	Pred. No. 2.6	
Matches	28	Conservative	17	Mismatches 38; Indels 25; Gaps 5.

[illegible]

RESULT	23
Q79076	Q79076
AC	Q79076;
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	VIF (VIRAL INFECTIVITY FACTOR).
CN	
OS	Human immunodeficiency virus type 1.
OC	Viruses; Retroloid viruses; Retroviridae; Lentivirus
NCBI	_taxid:_11676;

RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT A17.
RX MEDLINE-94303241: Pubmed-8030283;
RA Wieland U., Hartmann J., Suhr H., Salzberger B., Eggers H.J.,
RA Kuehn J.E.;
RT "In vivo genetic variability of the HIV-1 vif gene.";
RL Virology 203:43-51(1994).
DR EMBL: Z30684; CAAB3164.1; -;
DR InterPro: IPR000475; Viral_Infect.
DR Pfam: PF00559; Vif: 1.
DR PRINTS: PR00349; VIRIONINFECT.
DR PRODOM: PD000063; Viral_Infect; 1.
XQ SEQUENCE 192 AA: 22513 MW: 077054C1F87B21D8 CRC64;

Query Match	11.0%	Score 72.5	DB 12	Length 192
Best Local Similarly	25.6%	Pred. No. 2.6		
Matches 20	Conservative 12	Mismatches 37	Indels 9	Gaps 2

[illegible]

RESULT	24	
Q9ZHP0		
ID	Q9ZHP0	PRELIMINARY;
		PRT; 244 AA

DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, last annotation update)
DE NSP81 RESTRICTION ENDONUCLEASE.
GN NSP81R.
OS Nostoc sp. ATCC 29106.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_Taxid=82766;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC29106;
RX MEDLINE=99077292; PubMed=9862476;
RA Xu S.-Y., Xiao J.-P., Etwlinger L., Holden M., Aliotta J., Poh C.L.,
RA Dalton M., Robinson D.P., Petronzo T.R., Moran L., Ganatra M.,
RA Ware J., Slatko B., Benner J.,
RT *Cloning and expression of the ApuII, NspI, NspH1, SacI, ScaI, and
RT SapI restriction-modification systems in *Escherichia coli*.*,
RT Mol. Gen. Genet. 260:226-231(1998).
DR EMBL: AF056037; AAC97193.1; -.
DR REBASE: 1388; NspH1.
KW Endonuclease.
SQ SEQUENCE 244 AA; 27986 MW; B6A8C19A5D2A25B2 CRC64;

Query Match	11.0%	Score	72.5	DB	2	Length	244
Best Local Similarity	37.5%	Pred.No.	3.4				
Matches	21	Mismatches	6	Indels	18	Gaps	4

```

Oy 55 DSKLPAVQEMCAQEPVHLDSPA---IKHOF-LTGGDTQGRY-----CRSGLSTG 100
    | | | | | | | | | | | | | | | | | |
Db 132 DHFRP-MERMCASEPRLTSMDSDEIKQFKQLKKDASGNHLLKRSRCERCIKTG 186

```

RESULT	25	
Q73438		
ID	Q73438	
		PRELIMINARY;
		PRT; 104 AA

DT 01-NOV-1996 (TREMBLrel. 01, created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE VIRUS INFECTIVITY FACTOR.
GN VIF.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirae.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95191036; PubMed=7884906;
RA Sova P., Van Raust M., Gupta P., Balachandran R., Chao W., Itescu S.,
RA McKinley G., Volisky D.J.;
RT *Conservation of an intact human immunodeficiency virus type 1 vif
RT gene in vitro and in vivo.*
RL J. Virol. 69:2557-2564(1995).
DR EMBL: U42260; AAA83809.1; -;
DR Interpro: IPR000475; Viral_infect.
DR Pfam: PF00559; Vif. 1.
DR ProDom: PD000063; Viral_infect; 1.
SQ SEQUENCE 104 AA; 12534 MW; 2249B6ACA3C67412 CRC64;

Query Match	10.98;	Score 72;	DB 12;	Length 104;
Best Local Similarity	21.68;	Pred. No. 1.5;		
Matches	22;	Conservative	15;	Mismatches 41;
				Indels 24;
				Gaps 3;

```
QY      8 LLLNGV-----TWGPTVEAIFETQXSLMAESEHXLKTLGCDADADYGPFGDSRLPAV 61
      ::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      8 MIVWQYDQDRMRIRTWKSLVKHHIYISRKTKGVVYRIIHVESTNPRTSSVEHIIPLGDRLLIVT 67
```

QY 62 QEMGAQEPVHLDSPAIKHQFLLTGDTGRYKRCRSGTGWXO 103
 || | || | | | |
 DB 68 TYWG-----LNTGERD--WHLGQGSIDWRE 91

RESULT	26	
Q73410		
ID	Q73410	PRELIMINARY;
1	Q73410	PRT; 192 AA

DT	01-NOV-1996	(TREMblrel. 01, Created)
DT	01-NOV-1996	(TREMblrel. 01, Last sequence update)
DT	01-JUN-2001	(TREMblrel. 17, Last annotation update)
DE	VIRUS INECTIVITY FACTOR.	
DE	WITH	

OS Human immunodeficiency virus type 1,
OC Viruses; Retroid viruses; Retroviridae; Lentivirus
OX NCBI_TaxID=11676;
RN [1]

RX MEDLINE=95191036; PubMed=7884906;
 RA Sova P., van Raast M., Gupta P., Balachandran R., Chao W., Itescu S.,
 RA McKinley G., Volsky D. J.;
 RT "Conservation of an intact human immunodeficiency virus type 1 vif

RL J. Virol. 69:2557-2564(1995).
 DR EMBL; U41179; AAA83771.1; -
 DR InterPro; IPR000475; Viral_infect
 DR Pfam; PF00559; VIF; 1.

```
DR   Prodrom; PD000063; Viral_infect; 1.  
SQ   SEQUENCE 192 AA; 2265 MW; 4B157DA04D1B0ADD CRC64;
```

Query Match	10.9%	Score 72;	DB 12;	Length 192;
Best Local Similarity	21.0%	Pred. No. 3;		
Matches 21; Conservative	17;	Mismatches 38;	Indels 24;	Gaps 3;

```

      8 MIVQYDRMRIRIWKSLVKNHMYISKAKEMSYRHHYESTHPKISSEVOJPLGDARLYIT 67

```

68 TYWG-----LHTGERE--WHLGGVSIEM 89

RESULT	27	
Q73413		
ID	Q73413	PRELIMINARY;
		PRT; 192 AA

DT 01-NOV-1996 (TREMBLrel_01, Created)
DT 01-NOV-1996 (TREMBLrel_01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel_17, Last annotation update)
DE VIRUS INFECTIVITY FACTOR.
GN
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95191036; PubMed=7884906;
RA Sova P., Van Ranst M., Gupta P., Balachandran R., Chao W., Itescu S...
RA McKinley G., Volsky D.J.;
RT "Conservation of an intact human immunodeficiency virus type 1 vif gene in vitro and in vivo.";

RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paele J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.D., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Wortley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Rubin W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003603; AAF52055.1; -.
DR HSSP: P25779; IAIM.
DR IYBase: FBgn0037303; CG12163.
DR InterPro: IPR000010; Cystatin.
DR InterPro: IPR000668; Peptidase_C1.
DR InterPro: IPR003244; Sarcocystatin.
DR InterPro: IPR00169; Thiolprolactin.
DR Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS: PR00705; PAPA1N.
DR ProDom: PD031531; Sarcocystatin; 1.
DR SMART: SM00043; CY; 1.
DR PROSITE: PS00640; THIOI_PROTEASE_ASN; 1.
DR PROSITE: PS00139; THIOI_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOI_PROTEASE_HIS; 1.
KM Hydrolase: Thiol protease.
SQ SEQUENCE 475 AA; 53545 MW; 130DE83869498EF1 CRC64;

Query Match 10.9%; Score 72; DB 5; Length 475;
Best Local Similarity 24.8%; Pred. No. 8;
Matches 29; Conservative 22; Mismatches 38; Indels 28; Gaps 7;

OY 11 MGVY-WGPTFAIFETQXSLMAESENKLTGCGDADVPDPPGSRPAPVQEMQAE 69
DB 213 YGTFEADMTSS-EYKRTGLMORDE-AKATGSAAPVAYHGE-LPKEDMROKDA 266
OY 70 VHLDSFAIKHQ-----FLLTGDTQGRYRCRSGSLTGMXQLSKLLETGPVLC 118
DB 267 V-----TQVNNQSGSCGNAFSTGNTGIGLXAVKVG-----ELKPSSEQLLDC 310

RESULT 31
O9HUR4 PRELIMINARY: PRT; 489 AA.
AC O9HUR4:
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, last annotation update)
DE PROBABLE ALDEHYDE DEHYDROGENASE.
GN PA489.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=PA01:
RA MEDLINE-20437337; PubMed-10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong C.K.-S., Wu Z., Paulsen I.T.,
RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.:
RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004903; AAG08284.1; -.

DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyd; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 489 AA; 52697 MW; DECBAB6E6E73793D8 CRC64;

Query Match 10.8%; Score 71.5; DB 2; Length 489;
Best Local Similarity 27.4%; Pred. No. 9.4;
Matches 32; Conservative 16; Mismatches 54; Indels 15; Gaps 4;

OY 16 GPVTEAIFETQXSLMAESENKLTGCGDADVPDPPGDSRLPAPVQEMQAEVHLDS 74
DB 67 GFSARAAYLYKAVEVDKRRHEIVDWIIR-----ESGSTRKATIEGMAARAITLESA 119
OY 75 -PAIKHQLTGTQGR---YRCRSGSLTGMXQLSKLLETGPVLCALDGA 125
DB 120 SFPARVHGRIVESDVGKESRYRSAIGVAVISPNNFPLHLT-QRSIAPALALGNA 175

RESULT 32
O73412 PRELIMINARY: PRT; 192 AA.
AC O73412:
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, last annotation update)
DE VIRUS INFECTIVITY FACTOR.
GN VIF.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE-95191036; PubMed-7884906;
RA Sova P., van Ranst M., Gupta P., Balachandran R., Chao W., Itescu S.,
RA McKinley G., Volsky D.J.:
RT "Conservation of an intact human immunodeficiency virus type 1 vif
RT gene in vitro and in vivo.";
RL J. Virol. 69:2557-2564(1995).
DR EMBL: U4181; AAA83773.1; -.
DR InterPro: IPR00475; Viral_infect.
DR Pfam: PF00559; Vif; 1.
DR PRINTS: PR00349; VIRIONINFECT.
DR ProDom: PD000063; Viral_infect; 1.
SQ SEQUENCE 192 AA; 22707 MW; 5765EC32ACAE8E84 CRC64;

Query Match 10.7%; Score 71; DB 12; Length 192;
Best Local Similarity 21.0%; Pred. No. 3.9;
Matches 21; Conservative 16; Mismatches 39; Indels 24; Gaps 3;

OY 8 LILMGV-----TWGPTFAIFETQXSLMAESENKLTGCGDADVPDPPGDSRLPAPV 61
DB 8 MIYQVDRMRIRITWKSIVRHMHVSKAKEMSTRNHTESHPISSSEVOIPLGDARLVIT 67
OY 62 QEWGAQEPVHLDSPAIKHQLTGTQGRYRCRSGSLTGM 101
DB 68 TYWG-----LHTGERE-WHLGQGVSIEM 89

RESULT 33
O73435 PRELIMINARY: PRT; 192 AA.
AC O73435:
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, last annotation update)
DE VIRUS INFECTIVITY FACTOR.
GN VIF.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID-11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:95191036; PubMed:7884906;
 RA Sova P., van Ranst M., Gupta P., Balachandran R., Chao W., Ilescu S.,
 McKinley G., Volsky D.J.;
 RT "Conservation of an intact human immunodeficiency virus type 1 vif
 gene in vitro and in vivo."
 RL J. Virol. 69:2557-2564(1995).
 DR EMBL: U42257; AAA83806.1;
 DR InterPro: IPR000475; Viral_infect.
 DR Pfam: PF00559; Vif.1.
 DR PRINTS: PR00349; VIRIONINFCT.
 DR ProDom: PD000063; Viral_infect.1.
 SQ SEQUENCE 192 AA; 22357 MW; B7839EBA1A86E3D CRC64;

Query Match 10.7%; Score 71; DB 12; Length 192;
 Best Local Similarity 22.0%; Pred. No. 3.9;
 Matches 22; Conservative 14; Mismatches 40; Indels 24; Gaps 3;

QY 8 LLLMGV-----TWGPTVAIFETQXSLMAESEXLTLCGCCADVDVPGPDSRLPAV 61
 DB 8 MIVQVDRMRIRTKWKSIVKHHITISRTKGWVRHHIESTNPKTSSEVH1PLDARLVVT 67
 QY 62 QEWGAEVPHLDSPAIKHOFLLTGDYQGRYCRSGLTGW 101
 DB 68 TYWG-----LNTGERD-WHLGQVSIEM 89

RESULT 34
 ID 099BT8 PRELIMINARY; PRT; 192 AA.
 AC 099BT8;
 DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE VIF PROTEIN.
 GN VIF.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID-11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TV001-2;
 RA Scriba T.J., Treurnicht F.K., Zeier M., Engelbrecht S.,
 RA Tjane van Rensburg E.;
 RT "Characterization and phylogenetic analysis of South African HIV-1
 subtype C accessory genes."
 RL AIDS Res. Hum. Retroviruses 0:0-0(2001).
 DR EMBL: AF325742; AAK09096.1;
 SQ SEQUENCE 192 AA; 22527 MW; F8EDEB7598D78D05 CRC64;

Query Match 10.7%; Score 71; DB 12; Length 192;
 Best Local Similarity 25.8%; Pred. No. 3.9;
 Matches 17; Conservative 12; Mismatches 31; Indels 6; Gaps 1;

QY 8 LLLMGV-----TWGPTVAIFETQXSLMAESEXLTLCGCCADVDVPGPDSRLPAV 61
 DB 8 LIIWOVDRMKIRITWNSLVKHHMYISRRASGSYKHHFESRHPKVSSEVH1PLDARLVIK 67
 QY 62 QEWGAQ 67
 DB 68 TYWGLO 73

RESULT 35
 ID 046780 PRELIMINARY; PRT; 288 AA.
 AC 046780;
 DT 01-JUN-1998 (TReMBLrel. 06, Created)
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE MHC CLASS I HEAVY CHAIN (FRAGMENT).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID-9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BL;
 RA Urakawa T., Kodama M., Morita M., Ikeda H.;
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBD databases.
 CC -1 FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
 CC IMMUNE SYSTEM (BY SIMILARITY).
 CC -1 SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
 CC MICROGLOBULIN) (BY SIMILARITY).
 CC -1 SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 DR EMBL: AB008654; BAA24783.1;
 DR HSSP: P01900; 1B11.
 DR InterPro: IPR003597; Ig_C1.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR01039; MHC_I.
 DR Pfam: PF00047; Ig.1.
 DR Pfam: PF00129; MHC_I.1.
 DR ProDom: PD000050; MHC_I.1.
 DR SMART: SM00407; IGc1.1.
 DR PROSITE: PS00290; IG_MHC.1.
 KW Glycoprotein; MHC; Transmembrane.
 FT NON_TER 1
 FT NON_TER 288
 FT SEQUENCE 288 AA; 32994 MW; 833F22F170ECAB58 CRC64;

Query Match 10.7%; Score 71; DB 7; Length 288;
 Best Local Similarity 31.3%; Pred. No. 6;
 Matches 26; Conservative 13; Mismatches 28; Indels 16; Gaps 5;

QY 46 DADVPGPDSRLPAVOEWGAE-PVHLDSPAIKHOFLLTGDYQGRYCRSGLT--GM 101
 DB 35 DSDAPNPRKEPRAP-----WMEDGEPYWEA-----MTRDKKAKQQLRTGLNTIRGF 82
 QY 102 XOLSKLELTGPKVLACSLALDG 124
 DB 83 YNGSEAGSHTLQWLVLCGDVLDG 105

RESULT 36
 ID 045087 PRELIMINARY; PRT; 643 AA.
 AC 045087;
 DT 01-JUN-1998 (TReMBLrel. 06, Created)
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE C17H12.1 PROTEIN.
 GN C17H12.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Felodermidae; Caenorhabditis.
 OX NCBI_TaxID-6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE:94150718; PubMed-7906398;
 RA Wilson R., Alnscoough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton A., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL elegans.*
 Nature 368:32-38(1994).
 RL Nature 368:32-38(1994).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Geisel C.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RN [13]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF045642; AAC02580.1; -;
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 5.
 DR PRINTS: PR00320; GPROTEINRPT.
 DR SMART: SM00320; WD40; 5.
 DR PROSITE: PS50082; WD_REPEATS_2; 2.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD Repeat.
 SQ SEQUENCE 643 AA; 72048 MW; 987FC6D953BCB1E0 CRC64;

Query Match 10.6%; Score 71; DB 5; Length 643;
 Best Local Similarity 32.1%; Pred. No. 14;
 Matches 36; Conservative 17; Mismatches 31; Indels 28; Gaps 10;

QY 9 LLMGVTPTEAIFETQXSLMAESEHLKTLTGCDADVPGRPGDSRLPAVQEMGAQ- 67
 ID 533 LVLDAVMSVP-HPAVF-----ASIDADGNIFVWNLNE---DVEGV-ARLRA-----GAGG 578
 Db 578 ESYVMNRLIISPDKHVFV--GDDEGRVHMVDRESLYLKSSEMNFAFVL 628

RESULT 37
 ID 073436 PRELIMINARY; PRT; 147 AA.
 AC 073436;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE VIRUS INFECTIVITY FACTOR.
 GN VIF.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95191036; PubMed=7884906;
 RA Sovia P., van Ranst M., Gupta P., Balachandran R., Chao W., Itescu S.,
 RA McKinley G., Volisky D.J.;
 RT "Conservation of an intact human immunodeficiency virus type 1 vif
 RT gene in vitro and in vivo";
 RL J. Virol. 69:2557-2564(1995).
 DR EMBL: U42256; AAA83807.1; -;
 DR InterPro: IPR000475; Viral_Infect.
 DR Pfam: PF00559; Vif; 1.
 DR ProDom: PD000063; Viral_infect; 1.
 SQ SEQUENCE 147 AA; 17456 MW; AE0FC620FB3F9DE6 CRC64;

Query Match 10.6%; Score 70; DB 12; Length 147;
 Best Local Similarity 22.0%; Pred. No. 3.7;
 Matches 22; Conservative 14; Mismatches 40; Indels 24; Gaps 3;

QY 62 QEWGAQEPVHLDSPAIKHOFLTGTGGRYRCRSGLSTGW 101
 ID 68 TYWG-----LNTGERD--WHLGGVSIEM 89

RESULT 38
 ID 073434 PRELIMINARY; PRT; 192 AA.
 AC 073434;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE VIRUS INFECTIVITY FACTOR.
 GN VIF.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95191036; PubMed=7884906;
 RA Sovia P., van Ranst M., Gupta P., Balachandran R., Chao W., Itescu S.,
 RA McKinley G., Volisky D.J.;
 RT "Conservation of an intact human immunodeficiency virus type 1 vif
 RT gene in vitro and in vivo";
 RL J. Virol. 69:2557-2564(1995).
 DR EMBL: U42256; AAA83805.1; -;
 DR InterPro: IPR000475; Viral_infect.
 DR Pfam: PF00559; Vif; 1.
 DR PRINTS: PR00349; VIRIONINFECT.
 DR ProDom: PD000063; Viral_infect; 1.
 SQ SEQUENCE 192 AA; 22484 MW; 65C50FEB6D73FEC CRC64;

Query Match 10.6%; Score 70; DB 12; Length 192;
 Best Local Similarity 22.0%; Pred. No. 5;
 Matches 22; Conservative 14; Mismatches 40; Indels 24; Gaps 3;

QY 8 LLMGV-----TWGPTTEAIFETQXSLMAESEHLKTLTGCDADVPGRPGDSRLPAV 61
 ID 8 MIVQVDMRIRTWKSLVKHHIYISRTKGWYRRHHYESTNPKYSSEVHPLGDARLIVT 67
 Db 62 QEWGAQEPVHLDSPAIKHOFLTGTGGRYRCRSGLSTGW 101
 68 TYWG-----LNTGERD--WHLGGVSIEM 89

RESULT 39
 ID 09DPZ9 PRELIMINARY; PRT; 192 AA.
 AC 09DPZ9;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE VIF PROTEIN.
 GN VIF.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20569522; PubMed=11118362;
 RA Ngunj'u T., Renjifo B., Novitsky V.A., McLane M.F., GaoJekwe S.,
 RA Essex M.;
 RT "Molecular cloning and biological characterization of full-length HIV-
 RT 1 subtype C from Botswana";
 RL J. Virology 278:390-399(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ngunj'u T., Renjifo B., Novitsky V.A., McLane M.F., Essex M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF290030; AG34021.1; -;
 DR InterPro: IPR000475; Viral_infect.
 DR Pfam: PF00559; Vif; 1.

DR PRINTS; PRO0349; VIRIONINFECT.
 DR PRODOM; PD000063; Viral_Infect. 1.
 SQ SEQUENCE 192 AA; 22651 MW; 05E1B2B5508698C CRC64;

Query Match 10.6%; Score 70; DB 12; Length 192;
 Best Local Similarity 19.7%; Pred. No. 5;
 Matches 24; Conservative 19; Mismatches 43; Indels 36; Gaps 4;

OY 8 LLLMGV-----TWGPTEAIFETQXSLMAESEHXLKTLCGCCDADVPGPDSRLPAV 61
 DB 8 LIWQVDRMKIRITWNSLVKHHMTISKRANGFYRHHESHRPVSSEVMILPDCDARLVIT 67
 OY 62 QEWGAGEPVHLDSPAIRKHOFLLTGDTGGRYRCHSGLSTGWXQLSKLLETPKVLACSLA 121
 DB 68 TYWGLQ-----TGERE-WHLGHVSIEN-----RLRRCSTQ 97
 OY 122 LD 123
 DB 98 VD 99

RESULT 40
 O994S0 PRELIMINARY; PRT; 192 AA.

ID O994S0
 AC O994S0;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE VIF PROTEIN.
 GN VIF.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroviruses; Retroviridae; Lentivirus.
 OC NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=94IN476;
 RX MEDLINE=21094715; PubMed=11177395;
 RA Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
 Kallish M.L., Shaw G.M., Allen S., Hahn B.H., Gao F.;
 RT "Near full-length clones and reference sequences for subtype C
 isolates for HIV type 1 from three different continents."
 RL AIDS Res. Hum. Retroviruses 17:161-168(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=94IN476;
 RA Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
 Allen S., Shaw G.M., Hahn B.H., Gao F.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDAJ databases.
 DR EMBL; AF286223; AAK30956.1;
 SQ SEQUENCE 192 AA; 22625 MW; E0F8C15FE8211645 CRC64;

Query Match 10.6%; Score 70; DB 12; Length 192;
 Best Local Similarity 27.3%; Pred. No. 5;
 Matches 18; Conservative 10; Mismatches 32; Indels 6; Gaps 1;

OY 8 LLLMGV-----TWGPTEAIFETQXSLMAESEHXLKTLCGCCDADVPGPDSRLPAV 61
 DB 8 LIWQVDRMKIRITWNSLVKHHMTYSKRANGFYRHHESHRPVSSEVMILPDCDARLVIT 67
 OY 62 QEWGAG 67
 DB 68 TYWGLQ 73

RESULT 41
 O994S0 PRELIMINARY; PRT; 728 AA.
 ID O994S0
 AC O994S0;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, last annotation update)

DE COS54_22.
 GN COS54_22.
 OS Ranaid herpesvirus 1 (Lucke tumor herpesvirus).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.
 OC NCBI_TaxID=85655;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Davison A.J., Sauerbrier W., Dolan A., Addison C., McKinnell R.G.;
 RT "Genomic studies of the Lucke tumor herpesvirus (RaHV-1).";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DDAJ databases.
 DR EMBL; AF110004; AAD12285.1;
 SQ SEQUENCE 728 AA; 81366 MW; 78F7921CA573743D CRC64;

Query Match 10.5%; Score 69.5; DB 12; Length 728;
 Best Local Similarity 25.2%; Pred. No. 24;
 Matches 30; Conservative 11; Mismatches 59; Indels 19; Gaps 3;

OY 18 VTEAIFETQXSLMAESEHXLKTLCGCCDADVPGPDS-----RLPAQEWGAGE 68
 DB 250 VCAAGFYIIPVLPFAVERKXKHTSCFDALVQYVGERLQEDKQFLRLPAREMGTSF 309
 OY 69 PVHLDSPAIRKHOFLLTG-----DTGGRYRCHSGLSTGWXQLSKLLETPKVLACSLA 123
 DB 310 PALITCDVPRSYITCDPDSLDREWSLRTAQAQCCWN-----EVCVGRVVTNLILD 362

RESULT 42
 O54999 PRELIMINARY; PRT; 841 AA.

ID O54999
 AC O54999;
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE PAIRED-IG-LIKE RECEPTOR B (KILLER CELL INHIBITORY RECEPTOR-LIKE
 PROTEIN P91A).
 GN PIRB.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RA Takai T., Yamashita Y., Fukuta D., Tsuji A., Nagabukuro A.,
 RA Matsuda Y., Nishikawa Y., Ohyama Y., Ohmori H., Ono M.;
 RL J. Biochem. 125:0-0(1998).
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 DOMAIN.
 CC
 DR EMBL; AF040953; AAB97008.1; JOINED.
 DR EMBL; AF040946; AAB97008.1; JOINED.
 DR EMBL; AF040947; AAB97008.1; JOINED.
 DR EMBL; AF040948; AAB97008.1; JOINED.
 DR EMBL; AF040949; AAB97008.1; JOINED.
 DR EMBL; AF040950; AAB97008.1; JOINED.
 DR EMBL; AF040951; AAB97008.1; JOINED.
 DR EMBL; AF040952; AAB97008.1; JOINED.
 DR HSSP; P43626; INKR.
 DR MGD; MGI:894311; Pirb.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003600; Ig_Like.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig_5.
 DR SMART; SM00409; IG_1.
 DR SMART; SM00410; IG_Like_2.
 SQ SEQUENCE 841 AA; 93225 MW; 2E8E8FP97423385B CRC64;

Query Match 10.5%; Score 69.5; DB 11; Length 841;
 Best Local Similarity 27.2%; Pred. No. 28;
 Matches 31; Conservative 13; Mismatches 45; Indels 25; Gaps 5;
 OY 6 VFLLMGVTPGVTGA-----AIFETQXSLMAESEHXLKTLCGCCDADVPGPDSRLP 59

DB 215 VELLVSGNQLKPTIKAFESVITSKRAMTIMQ-----GLLDAEYVFLHNEGSGKQ 266
 QY 60 AVOEMCAOEPVHLDSPAIKHOFLLTGDTQ---GRYCRSGLSGTGWXQLSKLEL 110
 DB 267 STQT-----LQPGNKKEFFIPSMTRQHAGQYRCYCYGSAGWSQPSDLEL 312

RESULT 43

071262 PRELIMINARY; PRT; 192 AA.
 AC 071262;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE VIF PROTEIN.
 GN VIF.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_Taxid=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C18;
 RA Oelrichs R.B., McPhee D.A., Deacon N.J.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF042102; AAD03210.1; -;
 DR InterPro: IPR000475; Viral_infect.
 DR Pfam: PF00559; Vif; 1.
 DR PRINTS: PR00349; VIRIONINFECT.
 DR PRODOM: PD000063; Viral_infect; 1.
 SQ SEQUENCE 192 AA; 22500 MW; 22A97D251BC1ED40 CRC64;

Query Match 10.4%; Score 69; DB 12; Length 192;
 Best Local Similarity 20.6%; Pred. No. 6.4;
 Matches 21; Conservative 16; Mismatches 41; Indels 24; Gaps 3;

QY 8 LLLMGV-----TWGPVTEAIFETQXSLMASEHXLTLLGCGDADVPQPGDSRLPAY 61
 DB 8 MIWQVDRMRIRITWNSLVKHHVYSSKAKGMVIRHHENTHPRISSEVHIPLGDARLVYT 67
 QY 62 QEWGAOEPVHLDSPAIKHOFLLTGDTQGRYCRSGLSGTGXQ 103
 DB 68 TYWG-----LHTGERD--WHLGQVSIEMRE 91

RESULT 44

071966 PRELIMINARY; PRT; 192 AA.
 AC 071966;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE VIF PROTEIN.
 GN VIF.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_Taxid=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HIV-1JC;
 RA Mwaengo D.M., Novembre F.J.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF049494; AAC68843.1; -;
 DR InterPro: IPR000475; Viral_infect.
 DR Pfam: PF00559; Vif; 1.
 DR PRINTS: PR00349; VIRIONINFECT.
 DR PRODOM: PD000063; Viral_infect; 1.
 SQ SEQUENCE 192 AA; 22599 MW; 0D0E1456317A0673 CRC64;

Query Match 10.4%; Score 69; DB 12; Length 192;
 Best Local Similarity 20.2%; Pred. No. 6.4;

Matches 21; Conservative 18; Mismatches 41; Indels 24; Gaps 3;

QY 8 LLLMGV-----TWGPVTEAIFETQXSLMASEHXLTLLGCGDADVPQPGDSRLPAY 61
 DB 8 MIWQVDRMRIRITWNSLVKHHVYSSKAKGMVIRHHENTHPRISSEVHIPLGDARLVYT 67
 QY 62 QEWGAOEPVHLDSPAIKHOFLLTGDTQGRYCRSGLSGTGXQ 105
 DB 68 TYWG-----LHTGERD--WHLGQVSIEMRKRS 93

RESULT 45

071971 PRELIMINARY; PRT; 192 AA.
 ID 071971;
 AC 071971;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE VIF PROTEIN.
 GN VIF.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_Taxid=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HIV-1NC;
 RA Mwaengo D.M., Novembre F.J.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF049495; AAC68852.1; -;
 DR InterPro: IPR000475; Viral_infect.
 DR Pfam: PF00559; Vif; 1.
 DR PRINTS: PR00349; VIRIONINFECT.
 DR PRODOM: PD000063; Viral_infect; 1.
 SQ SEQUENCE 192 AA; 22398 MW; A6E790B042ABC996 CRC64;

Query Match 10.4%; Score 69; DB 12; Length 192;
 Best Local Similarity 20.2%; Pred. No. 6.4;
 Matches 21; Conservative 18; Mismatches 41; Indels 24; Gaps 3;

QY 8 LLLMGV-----TWGPVTEAIFETQXSLMASEHXLTLLGCGDADVPQPGDSRLPAY 61
 DB 8 MIWQVDRMRIRITWNSLVKHHVYSSKAKGMVIRHHENTHPRISSEVHIPLGDARLVYT 67
 QY 62 QEWGAOEPVHLDSPAIKHOFLLTGDTQGRYCRSGLSGTGXQ 105
 DB 68 TYWG-----LHTGERD--WHLGQVSIEMRKRS 93

Search completed: January 7, 2002, 16:51:26
 Job time: 273 sec

